IMPERIAL

DATA FOR POLICY .ORG

DATA & POLICY

CAMBRIDGE UNIVERSITY PRESS

8th International Conference

Data for Policy 2024

Decoding the Future: Trustworthy Governance *with* AI?

Book of Abstracts

9-11 July, 2024
Imperial College London, UK

Sponsors:

BILL & MELINDA GATES foundation

Microsoft

Data for Policy 2024 Book of Abstracts

Cite as:

Data for Policy CIC, **Engin**, Z.[1 2], **Kennedy**, M.[3], **Arcucci**, R.[3], **Crowcroft**, J.[4 5], & **Verhulst**, S.[6] (2024, July 24). Data for Policy 2024 Book of Abstracts. Data for Policy 2024, Imperial College, London. https://doi.org/10.5281/zenodo.12805534

Contact E-mail Address:

team@dataforpolicy.org

Published by **Data for Policy CIC**, London, United Kingdom.

[1] Data for Policy CIC
[2] University College London
[3] Imperial College London

[4] University of Cambridge
[5] The Alan Turing Institute
[6] New York University

Table of Contents

Conference Committees

General Chairs

Zeynep Engin, Data for Policy CIC, UK
Jon Crowcroft, University of Cambridge and The Alan Turing Institute, UK
Stefaan Verhulst, New York University, USA

Imperial Local Chairs

Rossella Arcucci, Imperial College London, UK
Mark Kennedy, Imperial College London, UK

Conference Management

Emily Gardner, Data for Policy CIC, UK
Pete Ford, Imperial College London, UK
Andrew Hyde, Cambridge University Press, UK
Pinar Ozgen, Data for Policy CIC, UK

International Committee

C. Leigh Anderson, University of Washington, USA
Maria Axente, PwC, UK
Emanuele Baldacci, European Commission, Luxembourg
Jennifer Hansen, Microsoft, USA
Innar Liiv, Tallinn University of Technology, Estonia
Samia Melhem, The World Bank, USA
Gianluca Misuraca, AI4Gov, Universidad Politécnica de Madrid, Spain
Francesco Mureddu, The Lisbon Council, Belgium
Peter Quartey, University of Ghana
Friederike Schueuer, UNICEF, USA
Roger Scott-Douglas, National Research Council of Canada
Barbara Ubaldi, OECD, France
Albrecht Wirthmann, European Commission
Masaru Yarime, The Hong Kong University of Science and Technology, Hong Kong

Imperial Local Committee

Anil A Bharath, Department of Bioengineering, Imperial College
Rafael A Calvo, Dyson School of Design Engineering, Imperial College
Yves-Alexandre de Montjoye, Department of Computing, Imperial College
David J Hand, Department of Mathematics, Imperial College
Mirko Kovac, Department of Aeronautics, Imperial College
Yingzhen Li, Department of Computing, Imperial College
Danilo Mandic, Department of Electrical and Electronic Engineering, Imperial College
Mirabelle Muuls, Imperial College Business School
Alessandra M Russo, Department of Computing, Imperial College
David L Shrier, Imperial College Business School
Kai Sun, Faculty of Engineering, Imperial College
Sanaz Talaifar, Imperial College Business School

Data for Policy Advisory Board

Kenneth Benoit, London School of Economics and Political Science, UK
Anil Bharath, Imperial College London, UK
Sir Anthony Finkelstein, City University, UK
Rayid Ghani, Carnegie Mellon University, USA
David Hand, Imperial College, UK
Lord Holmes of Richmond, UK All-Party Parliamentary Group on Data Analytics, UK
Christoph Lütge, Technical University of Munich, Germany
Helen Margetts, University of Oxford and The Alan Turing Institute, UK
Beth Noveck, New York University, USA
Alan Penn, University College London, UK
Rob Procter, University of Warwick and The Alan Turing Institute, UK
John Shawe-Taylor, UCL, UK
Peter Smith, University of Southampton, UK
Tom Smith, Dept of Levelling up, Housing & Communities, UK
John Taysom, Privitar, UK
Philip Treleaven, UCL, UK
Milan Vojnovic, London School of Economics and Political Science, UK
Dame Alison Wolf, King's College London, UK
Derek Wyatt, Digital Policy Alliance, UK

Area 1 Committee

Sarah Giest, University of Leiden, Netherlands
Sharique Manazir, Bharti Institute of Public Policy, India
Keegan McBride, Oxford Internet Institute, UK
Francesco Mureddu, Lisbon Council, Belgium
Anastasija Nikiforova, University of Tartu, Estonia
Sujit Sikder, Leibniz Institute of Ecological Urban and Regional Development, Germany

Area 2 Committee

Omar Isaac Asensio, Georgia Tech, USA
Giz Gulnerman, Ankara Hacı Bayram Veli Üniversitesi, Turkey
Cigdem Gurgur, Purdue University, USA
Jude Dzevela Kong, York University, Canada
Catherine Moore, Georgia Tech, USA
Emre Simsekler, Khalifa University of Science & Technology, UAE
Nicola Ulibarri, University of Calinfornia Irivine, USA

Area 3 Committee

Feras Batarseh, Virgina Tech, USA
Anushri Gupta, London School of Economics, UK
Sherman Kong, United Nations Foundation, Spain
Mihoko Sumida, Hitotsubashi University, Japan
Gaby Umbach, European University Institute, Italy
Laura Zoboli, University of Brescia, Italy

Area 4 Committee

Tristan Henderson, University of St Andrews, UK
Alexander Monea, George Mason University, USA
Mustafa Ozbilgin, Brunel University, UK
Jeannie Paterson, University of Melbourne, Australia
Nydia Remolina Leon, Singapore Management University, Singapore
Adrian Weller, The Alan Turing Institute, UK

Area 5 Committee

Bram Klievink, Leiden University, Netherlands
Itzelle Medina Perea, University of Sheffield, UK
Karen Yeung, University of Birmingham, UK
Leid Zejnilovic, Nova School of Business and Economics, Portugal

Area 6 Committee

Claire Boine, University of Ottawa, Canada
Eleonore Fournier-Tombs, United Nations University Centre for Policy Research, USA
Chris Hinnant, Florida State University, USA
Jim Jimeno, Office of the Undersecretary for Migrant Workers' Affairs, Philippines
Joni Jupesta, Research Institute of Innovative Technology for the Earth, Japan
Wilson Wong, Chinese University of Hong Kong, Hong Kong

Special Track Chairs

Stefaan Verhulst, The GovLab, New York University
Sara Marcucci, The GovLab, New York University
Martina Belmonte, Joint Research Centre of the European Commission
Alina Menocal Peters, IOM Global Migration Data Analysis Centre
Damien Jusselme, IOM Global Migration Data Analysis Centre
Matteo Fontana, Joint Research Centre of the European Commission
Anna Rosinska, Joint Research Centre of the European Commission
Rachel Adams, Global Center for AI Governance
Samuel Segun, Global Center for AI Governance
Julián Villodre, Leiden University
Sarah Giest, Leiden University
Victor Li, The University of Hong Kong
Jacqueline Lam, The University of Hong Kong
Jon Crowcroft, The University of Cambridge
Renan Gadoni Canaan, University of Ottawa
Teresa Scassa, University of Ottawa
Emanuele Baldacci, Eurostat
Ronald Jansen, United Nations
Albrecht Wirthmann, Eurostat

Conference Volunteers

Emily Eunji Kim, Georgia Institute of Technology
Huayizi Chen, UNICEF
Gemma E Ralton, Data Science Institute, Imperial College London
Georgia Meyer, LSE
Giz Gulnerman, Ankara HBV University
Gulsen Guler, Researcher and Data Literacy Consultant
Itzelle Medina Perea, University of Sheffield
Martin Gozzi, University College London
Muhammed Haider Faisan, Data for Policy CIC
Servet Yanatma, Data for Policy CIC
Ozlem Ayaz, Brunel University London
Diasmer Bloe, Researcher and Independent Consultant, Data for Policy CIC

Sponsors & Partners

Cambridge University Press
Imperial College London
Bill & Melinda Gates Foundation
Microsoft
UN Trade & Development (UNCTAD)
Smart Data Research UK
The Alan Turing Institute
Validate AI

Conference Programme (July 9th)

09:00 - 09:30 **Arrivals & Registration (Tea & Coffee)**

09:30 - 10:00 **Welcome & Opening Remarks**

Zeynep **Engin,** Chair & Director, Data for Policy CIC
Mark **Kennedy,** Director, Data Science Institute, Imperial College London
Professor Ian **Walmsley,** Provost of Imperial College London
"Our Nation in Numbers: The power of statistics in decision-making"
Professor Sir Ian **Diamond**, UK National Statistician

10:00 - 11:00 **Plenary 1**: **"Transforming Governance with AI & Trustworthiness"**

Speakers: Joel **Martin**, Chief Digital Research Officer & Chief Science Officer, National Research Council Canada / Gianluca **Misuraca**, AI4Gov Executive Director, Universidad Politécnica de Madrid, Spain / Barbara **Ubaldi**, Head of Digital Government and Data Unit, OECD / David **Shrier**, Professor of Practice, AI & Innovation, Imperial College London / Alexander **Iosad**, Senior Advisor, Government Innovation Policy, Tony Blair Institute

Chair: Zeynep **Engin,** Chair & Director, Data for Policy CIC (@dataforpolicy)

11:00 - 11:30 **Break**

11:30 - 12:30 **Parallel Session 1**

Session 1A : Exploring Open Data with Leaders in Industry

Speakers**:** Afua **van Haasteren**, Director, Health Policy & External Affairs, Roche / Carlos Martínez **Miguel**, Global Director – AI & Data Solutions and Services, Telefonica / Yiu-Shing **Pang**, Open Data Manager at UK Power Networks

Chair(s): Jennifer **Hansen**, Director, Open Data Policy & Strategy**,** Microsoft

Session 1B - St1.a: Digital & Data-driven Transformations in Governance

- *"How to design AI for public value: A socio-technical approach"* ; Viviana **Bastidas** - University of Cambridge,UK, Kwadwo **Oti-Sarpong** - University of Cambridge,UK, and Jennifer **Schooling** - Professor of Digital Innovation and Smart Places at Anglia Ruskin University. (*5049)*
- *"Investigating Public Sector Innovation Labs as-an-approach toward Data and AI-centric innovations in European National Governments*"; Francesco **Leoni** - Department of Design, Politecnico di Milano,Italy, Stefano **Maffei** - Department of Design, Politecnico di Milano,Italy and Bria **Jammali-Versace** - Department of Design, Politecnico di Milano,Italy. *(1039)*
- *"Data Governance in Data Altruism: archetypes definition";* Federico **Bartolomucci** - Politecnico di Milano,Italy, Edoardo **Ramalli** - Politecnico di Milano,Italy and Valeria Maria **Urbano** - Politecnico di Milano,Italy. *(1676)*
- *"Exploring the Intersection of Political Orientation and AI Governance Research: A Comprehensive Analysis of US Think-Tank Publications Using Large Language Models" ;* Emily Eunji **Kim** - Georgia Institute of Technology. *(4761)*

Chair: Keegan **McBride**, Oxford Internet Institute, University of Oxford

Session 1C - Sp2.a: Anticipating Migration for Policymaking

- *"Towards a Taxonomy of Anticipatory Methods: Integrating Traditional and Innovative Methods for Migration Policy" ;* Sara **Marcucci** -The Governance Lab, New York, United States of America and Stefaan **Verhulst** - The Governance Lab, New York, United States of America. (*DAP-2023-0173)*
- *"Augmentation or Replication? Assessing Big Data's Role in Migration Studies*" ; Tuba **Bircan** - Vrije Universiteit Brussel, Belgium. *(DAP-2023-0088)*
- *"Safeguarding migrant rights through open digital ecosystem (ODE) principles:a prerequisite for deploying anticipatory methods"* ; Rohan **Pai** - Aapti Institute and Amrita **Nanda** - Aapti Institute. *(7821)*
- *"Developing AI predictive migration tools to enhance humanitarian support. The case of EUMigraTool" ;* Cristina Blasi **Casagran** - Autonomous University of Barcelona, Barcelona, Spain, Mr Georgios **Stavropoulos** - Information Technologies Institute Centre for Research and Technology Hellas, Thessaloniki, Greece. *(DAP-2023-0095)*

Chair: Jon **Crowcroft,** University of Cambridge and The Alan Turing Institute

12:30 - 13:30 **Lunch**

13:30 - 14:30 **Keynote Lecture 1: "AI and Data Policy: Antagonism or Symbiosis?"**

Speaker: Jennifer **Prendki**, Head of Generative AI Data, Google DeepMind
Chair: Mark **Kennedy,** Director, Data Science Institute, Imperial College London

14:30 - 15:30 Parallel Session 2

Session 2A: Harnessing Data and AI for Climate Action: Bridging the Gap to Effective Policy

Speakers: Alyssa **Gilbert,** Director of Innovation at Grantham Institute for Climate change and the Environment, Imperial College London / Katharina **Weitz,** Project Manager & Researcher, Department of Artificial Intelligence, Fraunhofer HHI/ Massimo **Bonavita**, Principal Scientist, ECMWF - European Centre for Medium-Range Weather Forecasts / Robin **Lamboll**, Researcher, Imperial College London

Chair: Rossella **Arcucci**, Imperial College London

Session 2B: Making Digitalisation Environmentally Sustainable - organised in collaboration with **UNCTAD**

Speakers**:** Torbjörn **Fredriksson,** Head E-commerce and Digital Economy Branch, UN Trade & Development (presenting UNCTAD Digital Economy Report 2024) / Dorothea **Kleine**, Professor of Human Geography and Director of the Institute for Global Sustainable Development, University of Sheffield / George **Kamiya**, Independent Expert / Francesco **Mureddu,** Senior Director, The Lisbon Council

Chair: Zeynep **Engin**, Chair & Director, Data for Policy CIC

Session 2C: Developing Country Perspectives

- *"Knowledge, Attitudes and Readiness Towards Artificial Intelligence in Government Services; A developing Country Perspective" ;* Eric Afful-**Dadzie** - University of Ghana Business School and Samuel Lartey - University of Ghana Business School. *Sp3.(2122)*
- *"Towards A Fair and Equitable Data Ecosystem for Low Resource Languages" ;* Dorcas **Nyamwaya** - Equiano Institute,Nairobi, Kenya, Susan **Otieno** - Equiano Institute**,**Nairobi, Kenya, Chinasa **T. Okolo** - Equiano Institute, Nairobi, Kenya, Abigail **Oppong** - Equiano Institute, Nairobi, Kenya, and Jonas **Kgomo** - Equiano Institute,Nairobi, Kenya. *Sp3.(9826)*
- *"A political economy of information disorder in South and Southeast Asia" ;* Nicola **Nixon** - The Asia Foundation. *Sp4.(3059)*
- *"Social Media in Elections: A Glimpse of Mis/Disinformation from Developing Countries"*; Charmaine **Distor** - United Nations University, Danilo **Đikanović** - United Nations University and Soumaya **Ben Dhau -** United Nations University. *Sp4.(9103)*
- *"Community Based AI Governance"* ; Jonas **Kgomo** - Equiano Institute, Nairobi, Kenya. *(4596)*
  Chair: Stanley **Wood,** Evans School of Public Policy and Governance, University of Washington

Session 2D - St6.a: Global Challenges & Dynamic Threats

- *"AI and Digital Transformation of the Greater China Region: A Comparative Study of AI strategies in China, Taiwan, and Hong Kong" ;* Wilson **Wong** - The Chinese University of Hong Kong, Charles **Hinnant** - Florida State University, and Natalie **Wong** - National Chengchi University. *(465)*
- *"Blockchain as a new dynamic in interrogating AI-aided power centralisation with technological potential of decentralisation: A case of China's contested blockchain governance, and applying blockchain in data-driven governance" ;* Zichen **Hu** - London School of Economics and Political Science,UK. *(629)*
- *The 'Coup' Season: What Can Machine Learning and Artificial Intelligence Tell Us About the Resurgence of Coup D'Etats in Africa?" ;* Elikplimi K. **Agbloyor** - University of Ghana Business School, Boakye **Danquah** - University of Ghana Business School, Agyapomaa Gyeke **Dako** - University of Ghana Business School and Lei **Pan** - Curtin Univers*ity. (7406)*
  Chair: Roger Scott-**Douglas**, Acting President, National Research Council of Canada

15:30 - 15:50 Break

15:50 - 16:50 Parallel Session 3

Session 3A: Moving from Data Intelligence to Collective Decision Intelligence

Speakers: Sally **Cripps**, Human Technology Institute, University Technology, Sydney / Sir Geoff **Mulgan,** Professor of Collective Intelligence, Public Policy and Social Innovation, UCL **/** Stefaan **Verhulst,** The GovLab, New York University, Ben **Gales -** Chief Impact Officer, Paul Ramsay Foundation, and Gilad **Francis** - University of Technology, Sydney.

Chair: Alex **Fischer**, Human Technology Institute, University of Technology, Sydney

- "*Revisiting the assumptions around the data revolution as an accelerator of the Sustainable Development Goals*." ; Alex **Fischer** - Australian National University, Grant **Cameron** - United Nations Sustainable Development Solutions Network and Castelline **Tilus** - United Nations Sustainable Development Solutions Network. *7416 [3476]*
- *"Decision Accelerator Labs: Moving From Collective Data Intelligence to Collective Decision Intelligence."* ; Stefaan **Verhulst -** GovLab and Alex **Fischer** - Australian National University. *7416 [4388]*

- *"From data to systems intelligence" ;* Sir Geoff **Mulgan** - University College London. *7416 [2741])*
- *"BLADE: Bayesian Learning for Adversarial Defence" ;* Gilad **Francis** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia, Ngoc Lan Chi **Nguyen** - School of Computer Science, The University of Sydney, Sydney, NSW, Australia., Anna **Lopatnikova -** Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia */Discipline of Business Analytics, The University of Sydney, Darlington, NSW, Australia.*, Hadi Mohasel **Afshar** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia, Roman **Marchant** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia., Catarina **Moreira** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia. and Sally **Cripps** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia. *(DAP-2023-0178)*
- *"Bayesian Adaptive Trials for Social Policy"* ; Sally **Cripps** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia, Anna **Lopatnikova** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia. / Discipline of Business Analytics, The University of Sydney, Darlington, NSW, Australia. Hadi Mohasel **Afshar** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia., Ben **Gales** - Paul Ramsay Foundation, Darlinghurst, NSW, Australia, Roman **Marchant** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia, Gilad **Francis** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia, Catarina **Moreira** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia. and Alex **Fischer** - Australian National University, Canberra, ACT, Australia. *(DAP-2023-0179)*
- *"Bayesian Causal Discovery for Policy Decision Making" ;* Catarina **Moreira** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia, Ngoc Lan Chi **Nguyen** - School of Computer Science, The University of Sydney, Sydney, NSW, Australia, Gilad **Francis** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia, Hadi Mohasel **Afshar** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia, Anna **Lopatnikova** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia/Discipline of Business Analytics, The University of Sydney, Darlington, NSW, Australia, Sally **Cripps** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia, and Roman **Marchant** - Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia. *(DAP-2023-0180)*

Session 3B: Now you see me: Expert panel on data-driven decision making, insights from practice

Speakers: Alessandro **Paciaroni** - Research Associate, The Lisbon Council / Marcella **Bonanomi** - Senior Research Associate & Project Manager, Municipality of Milan / Antonio **Filograna** - Senior Researcher, Engineering Ingegneria Informatica S.p.A. *(6777)*

Chair: Francesco **Mureddu,** Senior Director, The Lisbon Council

Session 3C - St5.a: Algorithmic Governance

- *"Human-Machine Collaboration for Enhanced Decision-Making in Governance" ;* Dirk **Van Rooy** - University of Antwerp, Belgium. *(DAP-2023-0183)*
- *"Human oversight of algorithmic decisions: a post-deployment empirical investigation" ;* Susana **Lavado** - Nova School of Business and Economics, Charles **Wan** - Rotterdam School of Management, Erasmus University and Leid **Zejnilovic** - Nova School of Business and Economics. *(4411)*
- *"Trust in algorithmic governance: A meta-analysis" ;* Evrim **Tan** - KU Leuven. *(2715)*

Chair: Bram **Klievink,** Professor of Public Policy, Leiden University

Session 3D - St2.a: Technologies & Analytics

- *" "Smart or not": An Assessment Practice of Customer Service Chatbot from the Chinese Public Sector Based on Benchmark Testing" ;* Yuting **Huang** - School of Government, Peking University, Futian **Shao** - Laboratory for Government Big Data and Public Policy, Peking University and Weiyi **Zhang** - Global Development Institute, The University of Manchester. *(904)*
- *"Influence of Covid-19 Pandemic on Population-Level Behavioral Changes: An IoT Based Study in the USA"* ; Jasleen **Kaur** - University of Waterloo, Arlene **Oetomo** - University of Waterloo, Vivek **Chauhan** - University of Waterloo and Plinio **Morita** - University of Waterloo. *(6390)*
- *"AI for Women's Financial Inclusion – An Analysis of Product Design and Policy Approaches in Nigeria" ;* Adekemi Olufunmilola **Omotubora** - University of Lagos, Akoka, Nigeria. *(DAP-2023-0144)*

Chair: Omar Isaac **Asensio**, Director of the Data Science & Policy Lab, Georgia Institute of Technology

16:50 - 17:10	**Day 1 Closing Remarks: "The AI Revolution and London"**
	Speaker: Theo **Blackwell**, Chief Digital Officer for London, Mayor of London Chair: Zeynep **Engin,** Chair & Director, Data for Policy CIC
17:10	**End of Day 1 Programme**

Conference Programme (July 10th)

09:00 - 09:30 **Arrivals (Tea & Coffee)**

09:30 - 10:30 **Keynote Lecture 2 :**
"What we can learn from high-stake decisions in AI for medical treatment"

Speaker: Aldo **Faisal,** Director of Science & Innovation – Grand Challenges (Health), The Alan Turing Institute; Professor of AI & Neuroscience, Imperial College London
Chair: Stefaan **Verhulst,** The Gov Lab, New York University.

10:30 - 11:30 **Parallel Session 4**

Session 4A: At a Time of Rapid Advances in AI, Are We Instead Entering a Data Winter?

Speakers: Sonia **Cooper**, Open Innovation Team at Microsoft / Gina **Neff,** Executive Director of the Minderoo Centre for Technology & Democracy, University of Cambridge / Elena **Simperl**, Director of Research, Open Data Institute / Barbara **Ubaldi,** Head of Digital Government and Data Unit, OECD.

Chair: Stefaan **Verhulst**, The Gov Lab, New York University.

Session 4B - Sp3.a: AI, Ethics and Policy Governance in Africa

- *"Are Certain African Ethical Values at Risk from Artificial Intelligence?" ;* Samuel T. **Segun** - Global Center on AI Governance and The African Observatory on Responsible AI, South Africa. *(DAP-2023-0153)*
- *"Artificial Intelligence, Digital Colonialism and the Implications for Africa's Future Development" ;* Aishat **Salami** - Technology Consulting and Research, Veeta Advisory Hub, Lagos, Nigeria. *(DAP-2023-0174)*
- *"Should we communicate with the dead to assuage our grief? An Ubuntu perspective on governing griefbots" ;* Connor **Wright** - LCFI, University of Cambridge,UK, Montreal AI Ethics Institute, Montreal, Canada. *(DAP-2023-0141)*
- *"The ethics at the intersection of artificial intelligence and transhumanism: A personhood-based approach" ;* Amara Esther **Chimakonam** - Centre for Phenomenology in South Africa, University of Fort Hare, Alice, South Africa. *(DAP-2023-0132)*

Chair(s): Rachel **Adams -** Global Center on AI Governance and Samuel T. **Segun -** Global Center on AI Governance ; African Observatory on Responsible AI, South Africa

Session 4C - Sp6.a: Generative AI for Sound Decision-making

- *"Enhancing Health Policy-Making Through ChatGPT: Opportunities and Threats" ;* Shahabeddin **Abhari** - School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada, Plinio **Morita** - School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada and Jasleen **Kaur** - School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada. *(4258)*
- *"Generative AI and electric vehicle service operations in urban and remote areas" ;* Omar **Asensio** - Harvard Business School and Yifan **Liu -** Georgia Institute of Technology. *(9951)*
- *"Does Generative AI Revolutionize Higher Education? Perspectives, Policies, and Curriculum Reforms in Top Asian Universities" ;* Wilson **Wong** - The Chinese University of Hong Kong, Angela **Aristidou** - UCL, Konstantin **Scheuermann** - UCL and Tony **Wong** - The Chinese University of Hong Kong. *(8502)*

Chair: Victor **Li**, The University of Hong Kong

Session 4D - St3.a: Policy & Literacy for Data

- *"Commons for the Commons: Climate Action in the Amazon through Data Collaboratives" ;* Carolina **Banda** - Max Planck Institute for Innovation and Competition and University of Munich and Germán **Johannsen** - Max Planck Institute for Innovation and Competition and University of Munich. *(1340)*
- *"Leveraging data ecosystems to address climate challenges: an urban perspective" ;* Natalia **Oprea** - The Lisbon Council, SDA Bocconi School of Management, Charlotte **van Ooijen** - CvanO - Digital Government Research and Advice and Francesco **Mureddu** - The Lisbon Council. *(8347)*
- *"Datathon on Gender and Racial Inequalities in Public Service: an innovative data literacy experience from Brazil" ;* Carolina **Coppetti** - ENAP. *(7042)*

Chair(s): Ben **Snaith,** Open Data Institute

11:30 - 12:00 **Break**

12:00 - 13:00 **Parallel Session 5**

Session 5A: Better Together? How Smart Data and Public Data Can Co-Exist and Thrive - sponsored by the **ESRC Smart Data Research UK and the Alan Turing Institute**

Speakers: Joel **Suss**, Data Journalist, Financial Times / Martine **Wauben**, Head of Data for London, GLA / Anya **Skatova,** Senior Research Fellow, University of Bristol / Blair **Freebairn,** CEO, GEOLYTIX

Introducer**:** Mark **Birkin**, Director, Urban Analytics Programme, The Alan Turing Institute
Chair: Rachel **Franklin**, Newcastle University and The Alan Turing Institute

Session 5B - St1.b: Digital & Data-driven Transformations in Governance

- *"AI Product Cards: A framework for code-bound formal documentation cards in the public administration";* Albana **Celepija** - Fondazione Bruno Kessler, Trento, Italy, University of Trento, Trento, Italy, Alessio Palmero **Aprosio** - Fondazione Bruno Kessler, Trento, Italy, Bruno **Lepri** - Fondazione Bruno Kessler, Trento, Italy and Raman **Kazhamiakin** - Fondazione Bruno Kessler, Trento, Italy. *(DAP-2023-0165)*
- "AI documentation *method based on the Databook: case study of an audit of a fraud detection model";* Anna **Nesvijevskaia** - Conservatoire National des Arts et Métiers - DICEN Ile-de-France and Simon **Le Mouellic** - Quinten. *(8094)*
- *"Catching the bad apples to keep up the good work: city council perspectives on data-driven governance" ;* Margot **Kersing** - PhD, Erasmus University Rotterdam, Lieke **Oldenhof -** Associate professor, Erasmus University, Kim **Putters** - Professor, Tilburg University and Liesbet **van Zoonen** - Professor, Erasmus University. *(5934)*
- *"GRAIL: Developing responsible practices for AI and machine learning in research funding and evaluation with a community of learning" ;* Denis **Newman-Griffis** - University of Sheffield. *(4190)*
- *"Data-driven analysis of school performance measurement" ;* Ian **Widdows** - University of Sheffield. *(9082)*

Chair: Sarah **Giest ,** Professor of Public Policy, Leiden University

Session 5C - Sp7: Governance of Health Data for AI Innovation

- *"Signalling and rich trustworthiness in data-driven healthcare: an interdisciplinary approach" ;* Jonathan R **Goodman** - Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, UK and Richard **Milne** - Kavli Centre for Ethics, Science, and the Public, Faculty of Education, University of Cambridge, UK, Wellcome Connecting Science, Hinxton, Cambridge, UK. *(DAP-2023-0156)*
- *"Operationalizing health data governance for AI innovation in low-resource government health systems - a practical implementation perspective" ;* Tracey **Li** - D-tree, Zanzibar, Tanzania , Abbas **Wandella** - D-tree, Zanzibar, Tanzania, Richard **Gomer** - School of Electronics and Computer Science, University of Southampton, UK, and Mohamed Habib **Al-Mafazy** - Information and Communications Technology Unit, Ministry of Health, Zanzibar. *(DAP-2023-0157)*
- *"AI innovation in healthcare and platformization in Brazil: an analysis of the National Health Data Network (RNDS) under the right to health and personal data protection"* ; M. Matheus Zuliane **Falcão** - Centre for Law, Technology and Society, University of Ottawa, Ottawa, Canada, M. Raquel Requena **Rachid** - Oswaldo Cruz Foundation – Fiocruz, Rio de Janeiro, Brazil and Marcelo **Fornazin** - Oswaldo Cruz Foundation – Fiocruz, Rio de Janeiro, Brazil. *(DAP-2023-0176)*
- *"Assessing Health Equity in the IoT Era: A Study on Algorithmic Bias and Public Health Outcomes" ;* Thokozani **Hanjahanja-Phiri** - University of Waterloo, Jasleen **Kaur** - University of Waterloo, Arlene **Oetomo** - University of Waterloo and Plinio **Morita** - University of Waterloo*. (840)*

Chair(s): Renan Gadoni **Canaan,** University of Ottawa and Teresa **Scassa,** University of Ottawa

Session 5D - St6.b: Global Challenges & Dynamic Threats

- *"Resilience of territories in the face of hydrogeological risk: the role of mitigation interventions in Lombardy region" ;* Giovanni **Azzone** - Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Italy, Teresa **Bortolotti,** MOX - Department of Mathematics, Politecnico di Milano, Italy, Giulia **Piantoni -** Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Italy, Sara **Ratti** - Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Italy, and Piercesare **Secchi** - MOX - Department of Mathematics, Politecnico di Milano, Italy. *(432)*
- *"Future-Proofing Data Governance to Prepare for Climate Change" ;* Jacob **Leiken** - NYU School of Law, Beverley **Hatcher-Mbu** - Development Gateway: An IREX Venture, and Tom **Orrell** - Development Gateway: An IREX Venture. *(860)*
- *"An integrated decision support tool for assessing the risk of labour exploitation on fishing vessels" ;* Ruoyun **Hui** - Alan Turing Institute, Jamie **Hancock** - Alan Turing Institute, Jat **Singh** - University of Cambridge; Alan Turing Institute, Hannah **Thinyane** - Diginex, Mark **Briers** - Alan Turing Institute, and Anjali **Mazumder** - Alan Turing Institute*. (5990)*
- *"Understanding Discrepancies Between Self-reported and Measured Climate Shocks in Small-scale Agriculture" ;* Didier **Alia** - University of Washington, C. Leigh **Anderson** - University of Washington, Joaquin **Mayorga** - University of Washington, Rebecca **Toole -** University of Washington, Andrew **Tomes** - University of Washington and Stanley **Wood** - University of Washington. *(8944*)

Chair: Rossella **Arcucci**, Imperial College London

13:00 - 14:00 Lunch

14:00 - 14:30	**Policy Keynote : "Legislating for Ethical AI: the AI (Regulation) Bill"**
	Speaker: **Lord Holmes of Richmond,** House of Lords, UK Chair: Jon **Crowcroft,** University of Cambridge and The Alan Turing Institute
14:30 - 15:30	**Plenary Session 2: "Responsible AI for Decision-Making"**

Speakers: Merve **Hickok,** President at the Centre for AI and Digital Policy (CAIDP), Washington DC / Andrea **Renda,** Director of Research at the Centre for European Policy Studies (CEPS) / Maura **Grossman**, Research Professor, University of Waterloo, Canada *(Remote Participation)* / Masaru **Yarime**, Associate Professor, Division of Public Policy and Division of Environment and Sustainability, Hong Kong University of Science and Technology (HKUST).

Chair(s): Roger **Scott-Douglas,** Acting President, National Research Council of Canada and Mark **Kennedy,** Director, Data Science Institute, Imperial College London

15:30 - 16:00	**Break**
16:00 - 17:00	**Parallel Session 6**

Session 6A - Sp3.b: AI, Ethics and Policy Governance in Africa [2 online speakers]

- *"Responsible artificial intelligence in Africa: Towards policy learning" ;* Paul **Plantinga** - Human Sciences Research Council, South Africa, Kristophina **Shilongo** - Mozilla Foundation, Namibia, Oarabile **Mudongo -** Consumers International, Botswana, Angelique **Umubyeyi** - Independent, South Africa, Michael **Gastrow -** Human Sciences Research Council, South Africa and Gabriella **Razzano** - OpenUp, South Africa. *(DAP-2023-0133) (Remote Presentation)*
- *"Trust Norms for Generative AI Data Gathering in the African Context" ;* Abiola Joseph **Azeez -** Philosophy Department & Canadian Robotics and Artificial Intelligence Ethical Design Laboratory, University of Ottawa, Canada, and Tosin **Adeate** - Department of Philosophy, Olabisi Onabanjo University, Nigeria. (*DAP-2023-0169)*
- *"Case Studies of AI Policy Development in Africa" ;* Kadijatou **Diallo** - Harvard Kennedy School, Harvard University, Boston, Massachusetts, United States, Jonathan **Smith** - Meta, Menlo Park, California, United States , Chinasa T. **Okolo** - Center for Technology Innovation, The Brookings Institution, Washington D. C., United States, Dorcas **Nyamwaya** - Equiano Institute, Nairobi, Kenya, Jonas **Kgomo -** Equiano Institute, Nairobi, Kenya, and Richard **Ngamita -** Equiano Institute, Nairobi, Kenya. (*DAP-2023-0177)*
- *"Social Justice Considerations in Developing and Deploying AI in Africa" ;* Getachew Hailemariam **Mengesha -** School of Information Science, Addis Ababa University, Addis Ababa, Ethiopia, Elefelious Getachew **Belay -** School of Information Technology and Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia and Rachel **Adams** - Global Center on AI Governance. *(DAP-2023-0185) (Remote Presentation)*
  Chair(s): Rachel **Adams** - Global Center on AI Governance and Samuel T. **Segun** - Global Center on AI Governance and The African Observatory on Responsible AI, South Africa.

Session 6B - Sp10: AI and Data Science to Strengthen Official Statistics

- *"Measuring and reporting uncertainty of AI and machine learning tools in official statistics" ;* Violeta **Calian** - Statistics Iceland and Anton Örn **Karlsson** - Statistics Iceland. *(3843)*
- *"AI in German official statistics - from first steps to recent challenges" ;* Florian **Dumpert** - Federal Statistical Office of Germany. *(5483)*
- *"Harnessing Private Data for Public Policy: Organisational and Methodological Challenges, a focus on Mobile Phone and Card Transaction Data" ;* Marie-Pierre **Joubert** - INSEE, Latifa **Oukhellou** - COSYS-GRETTIA, Université Gustave Eiffel and David **Bounie** - Télécom Paris. *(4119)*
- "Advancing Public Diplomacy evaluations: AI and predictive *analytics to leverage the global power of Hallyu, The Korean Wave"*, Natalia **Grincheva** - The University of Melbourne and LASALLE College of the Arts, University of the Arts Singapore. *(9614) (Remote Presentation)*
  Chair: Joel **Martin,** Chief Digital Research Officer & Chief Science Officer, National Research Council Canada

Session 6C - St4.a: Ethics, Equity & Trustworthiness

- *"AI-assisted pre-screening of biomedical research proposals: ethical considerations and the pilot case of "la Caixa" Foundation" ;* Carla Carbonell **Cortés** - Area of Partnerships with Research and Health Institutions, "la Caixa" Foundation, Barcelona, Spain, César **Parra-Rojas** - SIRIS Lab, Research Division of SIRIS Academic, Barcelona, Spain, Albert **Pérez-Lozano** - Analytics & Artificial Intelligence, IThinkUPC S.L.U., Barcelona, Spain, Francesca **Arcara** - SIRIS Lab, Research Division of SIRIS Academic, Barcelona, Spain, Sarasuadi **Vargas-Sánchez** - SIRIS Lab, Research Division of SIRIS Academic, Barcelona, Spain, Raquel **Fernández-Montenegro** - Analytics & Artificial Intelligence, IThinkUPC S.L.U., Barcelona, Spain, David **Casado-Marín** - Area of Partnerships with Research and Health Institutions, "la Caixa" Foundation, Barcelona, Spain, Bernardo **Rondelli** - SIRIS Lab, Research Division of SIRIS Academic, Barcelona, Spain and Ignasi **López-Verdeguer** - Area of Partnerships with Research and Health Institutions, "la Caixa" Foundation, Barcelona, Spain. *(DAP-2023-0159)*
- *"The Dark Side of Large Language Models: Legal and Ethical Challenges from Stochastic Parrots and Hallucination" ;* Zihao **Li** - University of Glasgow & Stanford University. *(1471)*
- *"Responsible AI Mechanisms in Public Sector Organizations: A Realist Synthesis Review" ;* Ana **Gagua** - TU Delft, Technology, Policy, and Management faculty, Haiko **van der Voort**, TU Delft, Technology, Policy, and Management faculty, Nihit **Goyal,** TU Delft, Technology, Policy, and Management faculty, Alexander **Verbraeck,** TU Delft, Technology, Policy, and Management faculty. *(8637)*
- *"Towards Fairer AI: A Visual Synthesis of Bias Mitigation Tools and Training Frameworks*" ; Alenka **Guček** - Institut Jožef Stefan, Ljubljana, Slovenia, Tanja Zdolšek **Draksler -** Institut Jožef Stefan, Ljubljana, Slovenia, Matej **Kovacic -** Institut Jožef Stefan, Ljubljana, Slovenia**,** Andreas **Karabetian** - University of Piraeus, Pireas, Attica Greece, Konstantinos **Mavrogiorgos** -University of Piraeus, Pireas, Attica Greece**,** George **Manias -** University of Piraeus, Pireas, Attica Greece. *(DAP-2023-0161)*

  Chair: Nydia Remolina **Leon**, Singapore Management University, Singapore

Session 6D - Interwoven Realms: Data Governance as the Bedrock for AI Governance

Speakers: Friederike **Schüür**, Chief of Data Strategy and Data Governance, UNICEF / Silvana **Fumega,** Global Data Barometer / Marta **Poblet**, The Data Tank / Andrew **Dwyer,** Royal Holloway Research Portal

Chair: Stefaan **Verhulst,** The Gov Lab, New York University.

17:00	**Conference Group Photo**
17:10	**Conference Reception** (Sponsored by Cambridge University Press)
17:30	**Chairs' Remarks**
19:00	**End of Day 2 Programme**

Conference Programme (July 11th)

09:00 - 09:30 **Arrivals (Tea & Coffee)**

09:30 - 10:30 **Parallel Session 7**

Session 7A: "Delivering AI Assurance as a Service, Key Stakeholder Responsibilities" - organised in collaboration with **Validate AI and Imperial Data Science Institute**

Speakers: Charles **Kerrigan,** Partner at CMS Legal / Zeynep **Engin**, Chair & Director, Data for Policy CIC / Tirath **Virdee,** Data and AI Professional

Chair: Ed **Humpherson**, Director General, Office for Statistics Regulation (OSR), UK

Session 7B - St2.b: Technologies & Analytics

- *"Honest Computing: achieving demonstrable data lineage and provenance for driving data and process sensitive policies" ;* Florian **Guitton** - Data Science Institute, Imperial College London, London, United Kingdom, Axel **Oehmichen** - Data Science Institute, Imperial College London, London, United Kingdom, Secretarium Ltd, London, United Kingdom, Étienne **Bossé** - Secretarium Ltd, London, United Kingdom, and Yike **Guo** - Hong Kong University of Science and Technology,Clear Water Bay, Hong Kong. *(DAP-2023-0172)*
- *"Assessing Human Wellbeing in a Trustworthy AI World: The Complexity of Urban Data" ;* Ayse **Giz Gulnerman** - Land Registry and Cadastre Department, Ankara HBV University, Ankara, Türkiye, and Florian **Koch** - Department of Law and Economics, University of Applied Sciences HTW Berlin, Berlin, Germany. *(2684)*
- *"Predicting the Success of Mobile Money Retail Agents in Ghana: a comparative analysis of well-explored vs less-explored markets using Artificial Intelligence and Machine Learning" ;* Daniel **Osarfo** - University of Ghana, Peter **Quartey** - University of Ghana, Agyapomaa **Gyeke-Dako** - University of Ghana Business School and Elikplimi K **Agbloyor** - University of Ghana Business School. *(6720)*

  Chair: Anil A. **Bharath** - Imperial College London

Session 7C - St4.b: Ethics, Equity & Trustworthiness

- *"A Feminist Framework for Urban AI Governance: Addressing Challenges for Public-Private Partnerships" ;* Laine **McCrory** - Department of Communication and Culture, Toronto Metropolitan University, Toronto, Canada and Department of Communication and Media Studies, York University, Toronto, Canada. *(DAP-2023-0175)*
- *"Fair AI for All: Gender Equity in Decision-Making Algorithms in Sub-Saharan Africa" ;* Lilian Olivia **Orero** - SafeOnline Women Kenya SOW-Kenya. *(3577)*
- *"Gaming Data: Digital urban twins, open data platforms and the ethics surrounding governing data" ;* Fran **Meissner** - University of Twente, Florence **Chee** - Loyola University Chicago and Michael **Nagenborg** - University of Twente. *(1586)*
- *"Political misuses of biometric systems and the (re)production of power asymmetries" ;* Júlia **García-Puig** - Leiden University. *(1897)*

  Chair: Sanaz **Talaifar,** Imperial College London

10:30 - 10:45 **Break**

10:45 - 11:45 **Parallel Session 8**

Session 8A: Piercing the Veil: Technology's Role in Detecting Illegal Content

Speakers: Shubham **Jain,** PhD - Researcher at Imperial College London on online safety technologies and privacy / Andreas **Gutmann**, PhD - Senior Technologist Online Safety & Security at Ofcom / Rachel **Warner** - Barrister and former NCA investigator

Chair: Yves-Alexandre **de Montjoye**, Computational Privacy Group, Imperial College London

Session 8B - St3.b: Policy & Literacy for Data

- *"Identifying stakeholder motivations in normative AI innovation policy: a systematic literature review for research guidance" ;* Frederic **Heymans** - imec-SMIT, VUB, Brussels, Belgium, and Rob **Heyman** - imec-SMIT, VUB, Brussels, Belgium. *(DAP-2023-0163)*
- *"An Analysis of the Lifecycle of Generative Artificial Intelligence in Industrial Settings: Implications for Governing Risks and Responsibilities among Stakeholders" ;* Hillary **Giam** - The Hong Kong University of Science and Technology, and Masaru **Yarime** - The Hong Kong University of Science and Technology. *(9833)*
- *"Drafting an 'AI Policy' for organizational use: Development Gateway's experience" ;* Beverley **Hatcher-Mbu** - Development Gateway: An IREX Venture, Tom **Orrell -** Development Gateway: An IREX Venture, and Jacob **Leiken** - NYU Law School. *(4022)*

  Chair: Merve **Hickok,** President at the Centre for AI and Digital Policy (CAIDP), Washington DC

Session 8C - Sp2.b: Anticipating Migration for Policymaking

- "*Mobile phone data for anticipating displacements: Practices, opportunities, and challenges"*; Bilgeçağ **Aydoğdu** - Computing and Information Sciences, Utrecht University, Utrecht, Netherlands, Özge **Bilgili -** Interdisciplinary Social Science, Utrecht University, Utrecht, Netherlands**,** Suphi **Güneş -** Turkcell Technology, Istanbul, Turkey and Albert Ali **Salah** - Computing and Information Sciences, Utrecht University, Utrecht, Netherlands. *(DAP-2023-0084)*
- *"Mixed-frequency VAR: A new approach to forecasting migration in Europe using macroeconomic data" ;* Emily **Barker -** The University of Southampton, Jakub **Bijak -** The University of Southampton.*(DAP-2023-0108)*
- "Could we have seen it co*ming? Towards an early warning system for asylum applications in the EU" ;* Emily **Barker** - The University of Southampton, and Jakub **Bijak** - The University of Southampton. *(9304)*

  Chair: Innar **Liiv,** Tallinn University of Technology

Session 8D - St5.b: Algorithmic Governance

- "*Constituting an AI: Accountability Lessons from an LLM Experiment" ;* Kelsie **Nabben** - Max Weber Fellow, Robert Schuman Centre for Advanced Studies, European University Institute, Florence, Italy. *(DAP-2023-0123) (Remote Presentation)*
- *"Towards Algorithmic Accountability in the Public Sector" ;* Simone Maria **Parazzoli** - ISI Foundation. *(4001)*
- *"A systematic review of regulatory strategies and transparency mandates in AI regulation in Europe, the US, and Canada" ;* Mona **Sloane** - UVA School of Data Science, University of Virginia, Charlottesville VA, United States, and Elena **Wüllhorst** - King's College London. *(DAP-2023-0171)*
- *"How to construct a trustworthy AI ethical principle: Inspired by Feenberg" ;* Xiaomei **Wang** - Zhejiang University, and Huayu **Xin** - Zhejiang University and University of Edinburgh. *(7309)*

  Chair: Leid **Zejnilovic**, Nova School of Business and Economics

11:45 - 12:00	**Break**
12:00 - 13:00	**Keynote Lecture 3: "Expanding Academia's Role in Public Sector AI"**
	Speaker: Russell **Wald,** Deputy Director, Stanford Institute For Human-Centered Artificial Intelligence (HAI) Chair: Rossella **Arcucci**, Imperial College London
13:00 - 14:00	**Lunch**
14:00 - 15:00	**Parallel Session 9**

Session 9A: Historical Arc from WWII to AI: Collective Action for Global Equity

Speakers: Jude **Kong,** Dalla Lana School of Public Health, University of Toronto ; York University / Maximilian **Kasy,** Professor of Economics, University of Oxford *(Remote Participation)* / Rachel **Adams**, Global Center on AI Governance / Sanaz **Talaifar,** Imperial College London

Chair: Leigh **Anderson,** Marc Lindenberg Professor of Humanitarian Relief, International Development and Global Citizenship, University of Washington.

Session 9B - St3.c: The Future of Data Ownership and Sovereignty: An examination on current governance modalities and debate on anticipatory trends *(Area 3 - Designed Panel)*

Speakers: Fei **Liao** - Nanjing Audit University (*Remote Participation)* / Yaniv **Benhamou** - University of Geneva, Masaru **Yarime** - The Hong Kong University of Science and Technology / Ashraf **Shaharudin** - Department of Urbanism, Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, the Netherlands / Annelieke **van den Berg** - TNO

Chair: Johanna **Walker,** King's College London

- *"Open Data Commons Licenses and Collective Data Governance for Personal and Non Personal Data" ;* Yaniv **Benhamou** - University of Geneva, and Melanie **Dulong de Rosnay** - CNRS. *(4779)*
- *"Exploring Emerging Trends in Data Governance: An AI-Assisted Approach to Bibliometric and Text Analyses" ;* Mushan **Jin** - The Hong Kong University of Science and Technology, and Masaru **Yarime** - The Hong Kong University of Science and Technology. *(8454)*
- "The National Audit as a Tool of 'Governance by Data' in China: A Phenomenological Approach" ; Fei **Liao** - Nanjing Audit University, Mengjia **Gu** - Nanjing Audit University, Yi **Lu** - Nanjing Audit University and Shichao Zhou - Nanjing Audit University. *(6089) (Remote Presentation)*
- "Exploring the *Contributions of Open Data Intermediaries for a Sustainable Open Data Ecosystem" ;* Ashraf **Shaharudin** - Department of Urbanism, Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, the Netherlands, Bastiaan **van Loenen** - Department of Urbanism, Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, the Netherlands, and Marijn J**anssen** - Department of Engineering, Systems and Services, Faculty of Technology, Policy and Management, Delft University of Technology, Delft, the Netherlands. *(DAP-2023-0147)*
- *""Capabilities for governmental data ecosystems for solving societal challenges";* Annelieke **van den Berg** - TNO, Marissa **Hoekstra** - TNO and Anne Fleur **van Veenstra** - TNO. *(1320)*

Session 9C - Sp6.b: "Designing a Value-driven GAI Framework for Social Good: Embedding Social Good Values into GAI Models"

Speaker: Victor **Li,**The University of Hong Kong
Chair: Jacqueline **Lam**, The University of Hong Kong

- *"AI at the Bench: Legal and Ethical Challenges of Informing – or Misinforming – Judicial Decision-Making Through Generative AI" ;* David Uriel **Socol de la Osa** - Hitotsubashi University, Hitotsubashi Institute for Advanced Study, Graduate School of Law, Tokyo, Japan, and Nydia **Remolina** - Singapore Management University, Singapore; Fintech Track Lead, SMU Centre for AI and Data Governance, Singapore. *(DAP-2023-0148)*
- *"Risks and Best Practices for Using Generative AI in Judicial Decisions" ;* Yuya **Ishihara** - Hitotsubashi University, Faculty of Law, and Mihoko **Sumida** - Hitotsubashi Institute for Advanced Study .(*8339)*

15:00 - 15:15	**Break**
15:15 - 16:15	**Plenary Session 3: "The Global Challenge: Harnessing AI's Potential and Navigating Its Risks for a Better World"**

Speakers: Jennifer **Hansen,** Director of Open Data Policy & Strategy, Microsoft / Rossella **Arcucci,** Imperial College London / Friederike **Schüür,** Chief of Data Strategy and Data Governance, UNICEF

Chair: Jon **Crowcroft,** University of Cambridge and The Alan Turing Institute

16:15 - 16:45	**Reflections & Closing**

Conference Chairs' Reflections & Closing Discussion:
Mark **Kennedy**, Imperial College London / Rossella **Arcucci**, Imperial College London / Jon **Crowcroft,** University of Cambridge & The Alan Turing Institute / Stefaan **Verhulst**, The GovLab, New York University / Zeynep **Engin,** Data for Policy CIC.

16:45	**End of Data for Policy 2024**

Keynote Lecture 1

"AI and Data Policy: Antagonism or Symbiosis?"

Jennifer **Prendki***
Head of Generative AI Data, Google DeepMind

Chair: Mark **Kennedy,** Director, Data Science Institute, Imperial College London

Abstract

As Generative models improve and their applications become more pervasive in industry, it is hardly a surprise that training data is turning into a hot commodity for AI researchers. Naturally, governments and institutions have been responding to the growing demand with additional policies such as the EU AI Act in order to control potential adverse societal impact of the use of large-scale web data for AI development. But are data collection and processing policies unequivocally restraining AI research as often perceived by AI developers? In my talk, I will discuss how - if at all - data policies can be used as an accelerator to AI research, and how AI research itself can be leveraged to improve and optimize AI data policies.

* Jennifer Prendki is a 360 Data expert with 19 years of experience in Research and Industry with a track record of enabling Data for AI and AI for Data. Her personal mission is to prepare Society for the future. She is an AI innovator, a builder and a strategist. As a leader, she specializes in bootstrapping early-stage data and AI initiatives (especially under time and cash-constrained situations), and in steering dysfunctional ones. Being a role-model to the younger generation of women in Tech is a big part of her mission and she is passionate about attracting more women to careers in STEM.

Keynote Lecture 2

"What we can learn from high-stake decisions in AI for medical treatment"

Aldo **Faisal**[*]
Professor of AI & Neuroscience Department of Bioengineering - Faculty of Engineering,
Imperial College London

Chair: Stefaan **Verhulst**, The Gov Lab, New York University

[*] Professor Aldo Faisal is the Professor of AI & Neuroscience at the Dept. of Computing and the Dept. of Bioengineering at Imperial College London. He was awarded a prestigious UKRI Turing AI Fellowship (£2 Mio including industry partners). Aldo is the Founding Director of the £20Mio. UKRI Centre for Doctoral Training in AI for Healthcare that aims to transform AI for Healthcare research and pioneer training 100 PhD and Clinical PhD Fellows. He also holds a Chair in Digital Health at the University of Bayreuth (Germany).At his two departments, Aldo leads the Brain & Behaviour Lab focussing on AI & Neuroscience and the Behaviour Analytics Lab at the Data Science Institute. He is Associate Investigator at the MRC London Institute of Medical Sciences and is affiliated faculty at the Gatsby Computational Neuroscience Unit (University College London). He was the first elected Speaker of the Cross-Faculty Network in Artificial Intelligence representing AI in College on behalf of over 200 academic members.

Keynote Lecture 3

"Expanding Academia's Role in Public Sector"

Rusell **Wald***
Deputy Director, Stanford Institute For Human-Centered Artificial Intelligence (HAI)

Chair: Rossella **Arcucci,** Imperial College London

Abstract

AI has captured public attention and become a focal point for policymakers. Concerns about AI have evolved from niche academic discussions to widespread public discourse, influencing legislative actions worldwide. Currently, the focus is mainly on industry-driven AI products, sidelining the broader AI ecosystem and societal impacts. This industry-centric approach marginalizes academia and civil society, potentially skewing AI governance toward industry interests rather than public good. To address this, diverse stakeholder involvement is essential in AI development. Robust academic research is crucial for human-centered AI, driving scientific curiosity, training future AI leaders, and providing policymakers with an objective understanding of AI. This keynote will discuss how governments must boost investment in public sector AI research and propose policies to balance industry dominance with significant academic contributions.

* Russell Wald serves as the deputy director for the Stanford Institute for Human-Centered Artificial Intelligence (HAI). In this role he oversees HAI's research, education, communications, administrative activities, industry programs, and policy and society hub. Wald works with HAI's co-directors and faculty leaders to help shape the strategic vision and human-centered mission of HAI. He is a contributor to the NAIRR bill. He is part of a HAI seed grant research project titled, Addicted by Design: An Investigation of How AI-fueled Digital Media Platforms Contribute to Addictive Consumption. He is a member of AI Index Steering Committee and former term member of the Council on Foreign Relations and the Truman National Security Project.

Policy Keynotes

Professor Ian **Walmsley**
Provost of Imperial College London

"Our Nation in Numbers: The power of statistics in decision-making"

Professor Sir Ian **Diamond***
UK National Statistician

"The AI Revolution and London"

Theo **Blackwell***
Chief Digital Officer for London, Mayor of London

Chair: Zeynep **Engin,** Chair & Director, Data for Policy CIC

"Legislating for Ethical AI: the AI (Regulation) Bill"

Lord Holmes of Richmond
House of Lords, UK

Chair: Jon **Crowcroft,** University of Cambridge and The Alan Turing Institute

Abstract

AI Bill - Ethical AI - Lord Holmes of Richmond MBE (lordchrisholmes.com)

Plenary Session 1

Transforming Governance with AI & Trustworthiness

Speakers:

Joel **Martin**, Chief Digital Research Officer & Chief Science Officer, National Research Council Canada
Gianluca **Misuraca**, AI4Gov Executive Director, Universidad Politécnica de Madrid, Spain
Barbara **Ubaldi**, Head of Digital Government and Data Unit, OECD
David **Shrier**, Professor of Practice, AI & Innovation, Imperial College London
Alexander **Iosad,** Senior Advisor, Government Innovation Policy, Tony Blair Institute

Chair: Zeynep **Engin,** Chair & Director, Data for Policy CIC

Plenary Session 1 Description:

As AI weaves itself into the fabric of critical decision-making, we are witnessing fundamental shifts in governance. Radical transformations are afoot in the way democracies function, citizen services are delivered, and justice is delivered. This evolution sparks a fascinating debate. One side envisions a golden age where AI helps us tackle long standing problems – such as inequalities and environmental degradation – issues often rooted in ingrained human and institutional decision-making practices. Conversely, anxieties loom about a dystopian future where AI amplifies these very problems, jeopardising hard-won human rights and even our very autonomy and control over decision-making.

In the midst of this entangled discussion, governments worldwide find themselves in a pivotal struggle to adapt to the transformative power of AI while simultaneously imposing regulations to keep the technology in check. With the introduction of Large Language Models (LLMs) in particular, the global community has spent the last two years heavily discussing how to control this constantly evolving technology - arguably losing sight at times on the fundamental changes that are happening and/or can happen in collective governance processes.

This session intends to recalibrate the overall discourse around AI in governance, with AI increasingly coming into play to share decision-making power alongside humans and established institutions. It aims to clarify some of the language surrounding this transformation and explore what is on offer and what is at risk at a fundamental level in the public sector context. Understanding the core capabilities and behaviour of AI, alongside the dynamic nature of the public sector, is crucial for shaping a beneficial path for this technology. This session explores how to leverage AI's strengths to advance our collective decision-making processes in the following threads:

1) **Space of possibilities for AI in governance:** What are the key arguments for and against AI in public/collective decision-making processes? What are we striving for when we talk about governance with/by AI, and what are we trying to avoid at all cost? We will explore some of the most optimistic and pessimistic scenarios.

2) **Trust and autonomy in AI supported decision-making:** What fundamentally differentiates AI-supported decision-making? To what extent can 'artificial agency' be *integrated into* or *transform* existing human and institutional decision-making processes? How can we frame "trust" in AI-supported decision-making? Can we make AI sufficiently "human-centric"? Where do we place human "autonomy" in all this discussion?

3) **Dynamic contexts of AI in governance**: Governance discussions have a strong national sovereignty component, whereas the impact of cutting-edge AI technologies (such as ChatGPT) transcends borders. While policies of the profit-driven AI companies affect much larger constituencies, public mandate for policy-making still remains with governments that have limited capacity to leverage AI technology in the public interest. Sovereign priorities, interests, and cultures surrounding the use of AI in governance are also wide-ranging and sometimes conflict. The question arises: how can we navigate this complexity to set a beneficial course for this discussion?

This exploration of AI's potential in governance serves as a springboard for the next plenary session, "Responsible AI for Decision-Making." The broader conversation then culminates in the closing plenary session, "The Global Challenge: Transforming AI into a Force for Good."

Additional Resources:

- Trustworthy Governance *with* AI?, Introductory blog by Data for Policy 2024 Conference co-Chairs: Zeynep Engin, Stefaan Verhulst, David Hand, Jon Crowcroft, Mark Kennedy, Rossella Arcucci
- Governance of, by and with AI, keynote by Gianluca Misuraca at Data for Policy 2022 Conference - Brussels Edition.
- Governing in the Age of AI: A New Model to Transform the State, Tony Blair Institute Publication authored by Alexander Iosad, David Railton, and Tom Westgarth:
- Governing with Artificial Intelligence: Are governments Ready? OECD Artificial Intelligence Papers, June 2024, No. 20
- Sovereign GPTs: Aligning Values in AI for Development, UN Trade & Development (UNCTAD) blog by David L. Shrier and A. Aldo Faisal

Plenary Session 2

Responsible AI for Decision-Making

Speakers:

Merve **Hickok**, President at the Centre for AI and Digital Policy (CAIDP), Washington DC
Andrea **Renda**, Director of Research at the Centre for European Policy Studies (CEPS)
Maura **Grossman**, Research Professor, University of Waterloo, Canada *(Remote Participation)*
Masaru **Yarime**, Associate Professor, Division of Public Policy and Division of Environment and Sustainability, Hong Kong University of Science and Technology (HKUST).

Chair(s):

Roger Scott-**Douglas**, Acting President, National Research Council of Canada
Mark **Kennedy**, Director, Data Science Institute, Imperial College London

Plenary Session 2 Description:

As AI systems permeate critical decision-making across diverse sectors, including governments, the imperative for responsible and ethical development and deployment takes centre stage. This session delves into the multifaceted landscape of AI governance and regulation globally in critical decision-making contexts. Issues to be discussed include:

- The imperative for the public sector to have AI-specific procurement guidelines, to ensure AI and algorithmic decision-making systems respect human rights before they are embedded into the public infrastructure and have real-life consequences;
- How to implement the basic ethical principles for responsible AI, such as transparency, explainability, and inclusiveness, in practice by incorporating the socio-economic conditions of various countries and sectors, such as through hard law versus soft law and risk-based versus sector-based approaches; and
- Whether responsible AI ("RAI") is a well-understood concept that can be implemented given that we have no consensus on definitions, such as what it means for AI to be "unbiased" or "fair," and because attaining RAI entails trade-offs between important competing considerations that are presently technically impossible to achieve all at once, such as accuracy, transparency or explainability, privacy, and fairness, and for which there is also no consensus.

Plenary Session 3

The Global Challenge: Harnessing AI's Potential and Navigating Its Risks for a Better World

Speakers:

Jennifer **Hansen**, Director of Open Data Policy & Strategy, Microsoft
Rossella **Arcucci**, Imperial College London
Friederike **Schüür,** Chief of Data Strategy and Data Governance, UNICEF

Chair: Jon **Crowcroft,** University of Cambridge and The Alan Turing Institute

Plenary Session 3 Description:

The world faces complex problems – climate change, poverty, pandemics, major demographic shifts, political instabilities, a global debt crisis. Artificial intelligence (AI) holds significant promise in addressing these issues, offering innovative and effective solutions. However, AI development also brings ethical implications and potential risks that must be carefully managed. This panel will explore both the positive and negative impacts of AI, examining its applications in addressing society's challenges.

Panel: Special Session 1A

Exploring Open Data with Leaders in Industry

Speakers

Afua van **Haasteren**, Director, Health Policy & External Affairs, Roche
Carlos Martínez **Miguel,** Global Director – AI & Data Solutions and Services, Telefonica
Yiu-Shing **Pang**, Open Data Manager at UK Power Networks

Chair(s):

Jennifer **Hansen**, Director, Open Data Policy & Strategy, Microsoft

Panel Abstract

Industry leaders and experts will discuss the opportunities and challenges encountered by industries in making their data accessible. From data privacy, security risks, regulatory compliance, competitive implications to AI technologies, the panelists will share their perspectives, experiences, and strategies regarding the opening of private sector data for societal and business benefit.

Panel: Special Session 2A

Harnessing Data and AI for Climate Action: Bridging the Gap to Effective Policy

Speakers

Alyssa **Gilbert**, Director of Innovation at Grantham Institute for Climate change and the Environment, Imperial College London
Katharina **Weitz**, Department of Artificial Intelligence, Fraunhofer HHI
Massimo **Bonavita**, ECMWF - European Centre for Medium-Range Weather Forecasts
Robin **Lamboll,** Researcher, Imperial College London

Chair: Rossella **Arcucci,** Imperial College London

Panel Abstract

As the world grapples with the urgent challenges of climate change, the role of data and artificial intelligence (AI) in developing effective policies cannot be overstated. This panel aims to explore the intersection of data, AI, and climate, delving into the potential for technological innovations to drive sustainable solutions and inform policy decisions. AI technologies have a great potential to enhance the capability in Earth system observations and predictions. Few examples will be given in the area of data-driven weather forecasting, characterization of extreme events, enhancement of early warning systems and multi-data/multi-model fusion. AI capability to process different data modalities, images processing and text has the potential to offer more accurate and unbiased information for policy making and foster a more well-informed global community confronted by climate challenges. The introduction of AI technologies has the potential to transform standard practices in numerical weather prediction and climate prediction, as it has been the case in other fields of applied science. Will it be an evolution or a revolution? Are there limits to what data-driven technologies can do in our field? Debate is open.

Panel: Session 2B

Making Digitalisation Environmentally Sustainable

organised in collaboration with

UN Trade & Development (UNCTAD)

Speakers

Torbjörn **Fredriksson**, Head E-commerce and Digital Economy Branch, UN Trade & Development (presenting UNCTAD Digital Economy Report 2024)
Dorothea **Kleine,** Professor of Human Geography and Director of the Institute for Global Sustainable Development, University of Sheffield
George **Kamiya,** Independent Expert
Francesco **Mureddu**, Senior Director, The Lisbon Council

Chair: Zeynep **Engin,** Chair & Director, Data for Policy CIC

Panel Abstract

The relationship between digitalization and environmental sustainability is bidirectional. On the one hand, data-driven digital technologies, such as Internet of Things, cloud computing and artificial intelligence, can be powerful tools to fight environmental challenges, such as climate change. On the other hand, the fast expansion of digital use is leading to a growing environmental footprint.
The net impact of digitalization on environmental sustainability depends greatly on how we manage the direct and indirect effects associated with the production, use and disposal of various digital devices and ICT infrastructure. A continuation of the current digitalization trajectories is not consistent with the need to comply with the “planetary guardrails” related to climate, nature, soils and oceans. The topic is slowly gaining attention, but there is still a lack of robust information and research into the role of policy for achieving desirable outcomes.
This session discussed the environmental implications of the accelerating speed of digitalization and how it links to the concurrent transition towards a low-carbon economy. In this context, it also considered implications for countries at varying levels of development.
As highlighted in a new publication from the UN Trade and Development, the Digital Economy Report 2024, the production and use of digital devices, data centres and information and communications technology (ICT) networks account for an estimated 6% to 12% of global electricity use. Various studies suggest that the ICT sector emitted an estimated 0.69 to 1.6 gigatons of CO_2 equivalents in 2020, corresponding to 1.5% to 3.2% of global GHG emissions.

Worldwide, data centres are estimated to have consumed as much energy as France in 2022 – 460 terawatt-hours (TWh) of electricity. Their energy consumption is expected by the International Energy Agency to double to 1,000 TWh in 2026.
Digitalization's water consumption is also growing, which is cause for concern in a world where two billion people still lack access to safe drinking water. In the United States, one-fifth of data centre servers' direct water footprint reportedly comes from watersheds that are moderately to highly water-stressed.
Developing countries bear a disproportionate share of digitalization's ecological costs while reaping fewer benefits. For example, many of the strategic mineral resources needed for the digital transition are mined in developing countries, and significant amounts of waste related to digitalization is sent from developed to developing countries. Meanwhile, low-income countries are relatively poorly prepared for harnessing digital technologies for economic development and for mitigating environmental risk.
To tackle the environmental challenges of digitalization, the UN Trade and Development report argues that the world should transition to a more circular and inclusive digital economy. This will involve adopting sustainable practices throughout the entire lifecycle of digitalization – from design and production to usage and disposal – while ensuring equitable economic benefits.
The world is engaged in multiple discussions on how to achieve more sustainable outcomes from digitalization, such as through a Global Digital Compact and in the 20-year review of the World Summit of the Information Society. At the same time, various environmental process, such as those related to climate change, biodiversity and raw material depletion, are starting to give increased attention to the role of digitalization.
The session was co-organized by Dr. Zeynep **Engin,** one of the Data for Policy Conference Co-Chairs, and Mr. Torbjörn **Fredriksson**, Head of E-commerce and the Digital Economy, UN Trade and Development (UNCTAD). Following a presentation by Mr **Fredriksson,** views and comments were shared by Professor **Dorothea Kleine**, Director of the Institute for Global Sustainable Development at the University of Sheffield, Mr George **Kamiya,** an expert on the energy and climate impacts of digital technologies, and Dr. Francesco **Mureddu**, Senior Director at The Lisbon Council.

Key issues discussed included:

- What are the main environmental implications from digitalization?
- How will the growth of generative artificial intelligence impact on the environmental footprint of digitalization?
- What are good practices in terms of assessing the environmental impacts of digitalization?
- How can the academic community help governments, businesses and consumers to develop a stronger evidence base on which to take decisions.
- What are possible solutions that governments and other stakeholders can explore to foster more sustainable and inclusive outcomes from digitalisation?
- How can the voices and concerns of developing countries become better reflected as the world develops its responses to digitalization?

Panel: Session 3A

Moving from Data Intelligence to Collective Decision Intelligence

Speakers

Sally **Cripps**, Human Technology Institute, University Technology, Sydney
Sir Geoff **Mulgan,** Professor of Collective Intelligence, Public Policy and Social Innovation, UCL
Stefaan **Verhulst,** The GovLab, New York University
Ben **Gales -** Chief Impact Officer, Paul Ramsay Foundation
Gilad **Francis** - University of Technology, Sydney.

Chair: Alex **Fischer**, Human Technology Institute, University of Technology, Sydney

Sally **Cripps**[1], Stefaan **Verhulst** [2] and Alex **Fischer**[3]

1. *Human Technology Institute, University Technology, Sydney*
2. *The GovLab*
3. *Australian National University*

Sub. No: 7416

Panel Abstract

Data is proving critical to new pathways to inform decision making systems, along with associated (and emerging) phenomena such as advanced analytics, machine learning, and artificial intelligence. Yet while the importance of data intelligence for policymakers is now widely recognized, there remain multiple challenges to operationalizing that insight–i.e., to move from data intelligence to decision intelligence. This panel is an opportunity to explain what we mean by decision intelligence and discuss why it matters (abstract 1). The panel discusses how decision intelligence ensures that insights derived from data are more effectively integrated into decision-making processes (abstracts 2 and 6). The panel presents new advanced data analytic methods and state-of-the-art modelling technologies that are built for adaptive collective decision making (abstract 3 and 4). The panel includes discussions of how lived expertise and thick data is incorporated into these collective intelligence systems, and how diverse stakeholders are engaged to build that connective intelligence systems around specific overlapping decision nodes (abstract 4 and 5). The panel seeks to define the factors that enable collective decision intelligence and propose specific new approaches including Bayesian Adaptive trials as part of wider models for Decision Accelerator Labs.

Revisiting the assumptions around the data revolution as an accelerator of the Sustainable Development Goals.

Alex **Fischer**[1], Grant **Cameron**[2] - Castelline **Tilus**[2]

1. *Australian National University*
2. *United Nations Sustainable Development Solutions Network*

Sub. No: 7416 [3476]

Abstract

When the Sustainable Development Goals were negotiated in 2014, global policy makers assumed that the data revolution would significantly accelerate development outcomes by improving efficacy and efficiency of policy design to implementation, while also holding governments accountable. Seven years after adoption of the target-driven goals, progress towards the goals is reversing despite advances in statistical systems capability, the novel production of data and application of data science.

This article reconsiders the core assumptions for why the data revolution would drive and accelerate SDG progress, how data generates value for policy decision, and the contingencies between technology and human policy systems. The 2014 report to the United Nations Secretary General, "A World That Counts" (WTC) framed the data for development agenda and set out a series of recommendations. Within the agenda, there were a set of implicit assumptions driving the theory of change. This article assesses the explicit pathways outlined in the report: measuring for accountability; generation of more disaggregated and real-time data supplies and improve policymaking and implementation efficiency to achieve targets. By reviewing the subsequent experiences, we identify four core enabling pillars and six ways to adapt our assumptions to further drive these data-enabled pathways.

Our assessment suggests that while many of the ambitious recommendations of the WTC have been implemented at global levels to advance the production and use of data and statistics, they have not successfully impacted the SDGs outcomes. The assumptions that have been actioned include the technological progress to increase data collection and the SDGs frameworks as a standardizing force for data reporting and accountability.

Subsequent experience found the COVID-19 crisis as a catalyst for innovation and use of national statistical systems, not the SDGs. The focus on increased financing has not materialised for statistical systems, although planning and tracking of national systems have emerged and may have longer-term impacts.

In this panel discussion, we will propose four new assumptions: (1) that the value of data revolution is contingent upon the policy formation processes and role of politics within decision-making systems; (2) impact of data and statistics is shaped by the

trustworthiness of our information systems; (3) capability to interpret and apply data is an equal gap to finance in terms of generating value; and (4) not all data have equal value for SDG pathways and thus more nuanced approach is required to collect the data what matters most to accelerate progress.
This panel, with a forthcoming paper to be submitted separately, has significant policy implications. It suggests that the advocacy for more financing should be targeted and shaped by assessments of the value of data. The discussion will explore how further innovation is emerging to challenge the core assumptions driving the global data revolution, including use of collective decision intelligence and Bayesian adaptive trials to identify what data matters most to advance the policy challenge.

Decision Accelerator Labs: Moving From Collective Data Intelligence to Collective Decision Intelligence.

Stefaan **Verhulst**[1] and Alex **Fischer**[2]

1. *GovLab*
2. *Australian National University*

Sub. No: 7416 [4388]

Abstract

We live at a moment of perhaps-unprecedented global upheaval. From climate change to pandemics, from war to political disharmony, misinformation, and growing social inequality: policy and social change makers today face not only new challenges, but new types of challenges. In our increasingly complex and interconnected world, existing systems and institutions of governance, marked by hierarchical decision making, are increasingly being replaced by overlapping nodes of multi-sector decision making. Information is increasingly politicized and fragmented.

Data is proving critical to these new forms of decision making, along with associated (and emerging) phenomena such as Bayesian analytics, machine learning, and artificial intelligence. Yet while the importance of data intelligence for policymakers is now widely recognized, there remain multiple challenges to operationalizing that insight–i.e., to move from data intelligence to decision intelligence.

This paper identifies six obstacles that prevent policymakers and others from translating insights into action. These challenges include lack of awareness of data's potential; poor problem definition; lack of technical capacity; lack of advanced methods; limited inclusiveness; and fragmented approaches. The paper will present a new model to overcome these challenges: the creation of decision accelerator labs. These labs, operating on a hub and spoke model, offer a collective intelligence platform to facilitate the development of evidence-based, targeted solutions to public problems and dilemmas. Broadly, collective decision intelligence focuses on the approach to leverage the greatest value from data into polycentric decision systems. They start by addressing four critical challenges:

1. Insights derived from data are not consistently or effectively integrated into decision-making processes, or providing value to the points where there are collective decision needs. This can lead to collecting data that does not support decisions and missing the data that matters most or being biased without sufficient situational awareness.
2. Decision makers are not effectively leveraging advanced data analytical methods and state-of-the-art technologies including simulations and uncertainty sciences.

3. Lived expertise, and thick data, are not sufficiently or effectively integrated into the decision-making process, resulting in less empathetic and less effective policies and interventions.
4. Collaboration between diverse stakeholders are not incentivised and thus silos are sustained. This often results in limited data sharing and barriers to open data access to data (when appropriate).

The above challenges represent real obstacles to fulfilling the potential of data in decision making by policymakers and others. Some possible solutions have been mentioned–e.g., greater use of data collaboratives to overcome data silos, a new science of questions to help establish priorities. In this section, we propose the greater use of an innovative institutional structure that we call Decision Accelerator Labs (DALs).

DALs are a new institutional structure that function as connectors, bringing together different stakeholders that play a role in the decision making process. These stakeholders can include parties that need to be consulted or informed (e.g., citizens); parties who play a role as validators (e.g., domain experts or those with lived experience); translators (who translate information or data into meaningful action); and of course decisionmakers themselves.

DALs can take many forms. They could for instance take the shape of multi-party Bayesian adaptive trials challenges that enable communities and analysts to inform program experimentation; as think tanks that bridge sectors in training and analytics; or perhaps virtual immersive and interactive decision theaters that provide a conducive environment for decision makers to visualize data, models, and scenarios. In all cases, DALs should follow a flexible and adaptable spoke-and-hub model, allowing for the creation of tailored decision-making environments that can address diverse challenges across different scales, sectors, and issues (e.g., at the global or regional level, or in a particular domain). By enabling DALs to cater to diverse contexts, the spoke-and-hub model thus ensures that stakeholders have access to the most relevant and effective resources, fostering informed decision-making and enabling more targeted solutions.

DALs would help decision makers (as well as those affected by their decisions) in several ways. Some of their key enhancements would include:

- Question science: By bringing together experts and other stakeholders, DALs will play a key role in advancing participatory question science to ensure decision makers are asking the questions that matter most. In this exercise, DALs can leverage and build upon The GovLab's 100 Questions experience and methods.
- Decision mapping and requirements: DALs can develop new methods to identify decision requirements, as well as where there is greatest need for decision support and what tools or systems could offer such support. This process involves mapping existing decision systems and needs to identify priorities.
- Leverage Advanced Data Analytical Methods: DALs will employ new data analytical techniques, such as machine learning, artificial intelligence, advanced simulation, and network analysis to uncover hidden patterns and relationships within the data. These innovative methods will help identify relevant factors and variables, enabling more accurate predictions and actionable insights; in so doing, they will lead to more effective and responsive policies, and help define the boundaries of 21st century decision making.

- Integration of Lived Experiences and Thick Data: Recognizing the value of firsthand accounts and real-time data, DALs can help incorporate lived experiences and thick data (granular, high-frequency data) into decision-making processes. It will enable such processes through robust thick data-driven methods such as digital ethnography combined with big data-driven insights. This approach will ensure that decisions and policies that emerge from the DALs reflect ground realities and contribute to more inclusive, empathetic, and effective solutions.
- Rapid Deliberation and Iteration: Relying on real-time data, advanced simulation and analytics, and other methods, DALs will incorporate feedback from experts and policymakers to improve upon its outputs and models iteratively and quickly. In addition, collective intelligence methods and tools can be used to evaluate the impact of policies and programs and test ideas through simulations.

The paper will expand the evidence base sitting behind each of these components, including a review of current methods, applications, and suggest options for potential future design of Decision Accelerator Labs.

From data to systems intelligence.

Sir Geoff **Mulgan**
University College London

Sub. No: 7416 [2741]

Abstract

The paradox of our times is that we have abundant data and information but so often not much sign of more intelligent decisions and outcomes. Data has come to play a vital role in analysis, diagnosis of problems and in implementation of policies, but on its own its never enough. Here I suggest some ways in which the field needs to evolve in the next decade.

Intelligence-based organisation. The first is better organisation. Intelligence is not costless. It requires hard work to orchestrate, curate, analyse, link and interpret. The pandemic showed that the intelligence function needs to become central to the work of governments. It is currently divided not just by functional silos (health, economics, security etc) but also by professional silos (data, evidence, statistics, policy). Yet the full benefits depend on integration.

Seeing intelligence as an outcome not just an input. The great majority of discussion focuses on specific tools and inputs – open data, AI, evidence synthesis. Yet these are only as useful as the intelligent decisions that result. I argue for slipping models on their head, starting with desired outcomes and working backwards to mobilise the many sources of intelligence that can contribute. This can then guide improvements in many of the specific elements of intelligence, including continuous mobilisation of evidence, making use of the constantly evolving generative AI tools; being smart about transferability (what works when, where, how and why?), improving tools such as systems maps; and much more attention to synthesis – a gap in methods and organisation which so far AI cannot help with much.

New institutions. Finally, I argue that we need a new generation of public institutions many of which will need data and intelligence at their core, whether for care or mental health, energy transitions or the protection of democracy. One group will need to orchestrate and curate data for the public benefit. There are some promising examples, from India and Estonia to Taiwan. Yet these remain missing in most of the world and for most of the priorities of the next decade, with data largely proprietary, opaque and unlinked, making it impossible to train and mobilise AI.

BLADE: Bayesian Learning for Adversarial Defence

Gilad **Francis**[1], Ngoc Lan Chi **Nguyen**[2], Anna **Lopatnikova**[1,3], Hadi Mohasel **Afshar**[1], Roman **Marchant**[1], Catarina **Moreira**[1] and Sally **Cripps**[1]

1. *Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia*.
2. *School of Computer Science, The University of Sydney, Sydney, NSW, Australia*.
3. *Discipline of Business Analytics, The University of Sydney, Darlington, NSW, Australia*.

Sub. No: DAP-2023-0178

Abstract

Governments interested in de-escalating violent conflicts must infer and understand the complex interaction between international actors. This study investigates the application of Bayesian causal discovery as a method for conflict modeling utilizing publicly accessible event data. To illustrate its effectiveness, we conducted a case study on the conflict between Sri Lanka and the Tamils from 2000 to 2011. Employing a Bayesian graphical model, we aim to model conflict data and unravel the interconnected pathways between the actions of opposing parties.

The analysis brings to light challenges decision-makers and practitioners may encounter while modeling and examining societal events, such as data integrity, modeling deficiencies and analysis interpretation.

Bayesian Adaptive Trials for Social Policy

Sally **Cripps**[1], Anna **Lopatnikova**[1,2], Hadi Mohasel **Afshar**[1], Ben **Gales**[3], Roman **Marchant**[1], Gilad **Francis**[1], Catarina **Moreira**[1] and Alex **Fischer**[4]

1. *Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia*.
2. *Discipline of Business Analytics, The University of Sydney, Darlington, NSW, Australia*.
3. *Paul Ramsay Foundation, Darlinghurst, NSW, Australia*.
4. *Australian National University, Canberra, ACT, Australia*.

Sub. No: DAP-2023-0179

Abstract

This paper proposes Bayesian Adaptive Trials (BAT) as both an efficient method to conduct trials and a unifying framework for evaluating social policy interventions, addressing limitations inherent in traditional methods such as Randomized Controlled Trials (RCT). Recognizing the crucial need for evidence-based approaches in public policy, the proposal aims to lower barriers to the adoption of evidence-based methods and align evaluation processes more closely with the dynamic nature of policy cycles. BATs, grounded in decision theory, offer a

dynamic, “learning as we go” approach, enabling the integration of diverse information types and facilitating a continuous, iterative process of policy evaluation. BATs’ adaptive nature is particularly advantageous in policy settings, allowing for more timely and context-sensitive decisions. Moreover, BATs’ ability to value potential future information sources positions it as an optimal strategy for sequential data acquisition during policy implementation. While acknowledging the assumptions and models intrinsic to BATs, such as prior distributions and likelihood functions, the paper argues that these are advantageous for decision-makers in social policy, effectively merging the best features of various methodologies.

Bayesian Causal Discovery for Policy Decision Making

Catarina **Moreira**[1], Ngoc Lan Chi **Nguyen**[2], Gilad **Francis**[1], Hadi Mohasel **Afshar**[1], Anna **Lopatnikova**[1,3], Sally **Cripps**[1] and Roman **Marchant**[1]

1. *Human Technology Institute, University Technology Sydney, Ultimo, NSW, Australia*.
2. *School of Computer Science, The University of Sydney, Sydney, NSW, Australia*.
3. *Discipline of Business Analytics, The University of Sydney, Darlington, NSW, Australia*.

Sub. No: DAP-2023-0180

Abstract

This paper demonstrates how learning the structure of a Bayesian network, often used to predict and represent causal pathways, can be used to inform policy decision-making. We show that Bayesian Networks are a rigorous and interpretable representation of interconnected factors that affect the complex environment in which policy decisions are made. Furthermore, Bayesian structure learning differentiates between proximal or immediate factors and upstream or root causes, offering a comprehensive set of potential causal pathways leading to specific outcomes.

We show how these causal pathways can provide critical insights into the impact of a policy intervention on an outcome. Central to our approach is the integration of causal discovery within a Bayesian framework, which considers the relative likelihood of possible causal pathways rather than only the most probable pathway.

We argue this is an essential part of causal discovery in policy making because the complexity of the decision landscape inevitably means that there are many near equally probable causal pathways. While this methodology is broadly applicable across various policy domains, we demonstrate its value within the context of educational policy in Australia. Here, we identify pathways influencing educational outcomes, such as student attendance, and examine the effects of social disadvantage on these pathways. We demonstrate the methodology's performance using synthetic data and its usefulness by applying it to real-world data. Our findings in the real example highlight the usefulness of Bayesian networks as a policy decision tool and show how data science techniques can be used for practical policy development.

Panel: Session 3B

Now you see me: Expert panel on data-driven decision making, insights from practice

Speakers
Alessandro **Paciaroni -** Research Associate, The Lisbon Council
Marcella **Bonanomi** - Senior Research Associate & Project Manager, Municipality of Milan
Antonio **Filograna** - Senior Researcher, Engineering Ingegneria Informatica S.p.A.

Chair: Francesco **Mureddu**, Senior Director, The Lisbon Council

Sub. No: 6777

Panel Abstract
As technology plays an increasingly integral role in gathering, processing, and presenting information that drives decision making, understanding human-computer interaction is crucial. Effective human-computer interaction can significantly enhance the accuracy, efficiency and outcomes of decisions, particularly in high-stakes environments such as emergency response and public policy formulation. By examining how individuals interact with data and technology, we can identify ways to optimise these processes, reduce errors, and improve overall decision quality.

*How does the **human interact** with the **technology** and the **data** during the **decision making** process involved in public service provision of emergency services and in policy making?*

The expert panel will focus on sharing insights and results from the experts and protagonists of different Horizon projects. Key questions addressed in the conversation will include the differences and similarities in these interactions, the role of timeliness and emotional responses, and how these factors shape behaviour and outcomes. This expert panel will feature insights from leading digital transformation companies collaborating with large public sector institutions, innovative SMEs working closely with first responders, local governments, and think tanks. By contrasting and comparing user experience design and human-computer interaction across diverse projects, the audience will hear directly from the experience of large-scale research and innovation projects developing, testing and deploying solutions for data-driven decision making in the public sector. The use cases range from greenery management in urban areas to policy formulation for emergency preparedness and first responders operations management and training in the field. Attendees will gain valuable perspectives on optimising data and technology use to enhance decision-making and improve public service delivery. Join us to

learn from leading experts and engage in a critical discussion on advancing public service through better human-computer interaction.

Panel: Session 4A

At a Time of Rapid Advances in AI, Are We Instead Entering a Data Winter?

Speakers

Sonia **Cooper,** Open Innovation Team at Microsoft
Gina **Neff,** Executive Director of the Minderoo Centre for Technology & Democracy, University of Cambridge
Elena **Simperl,** Director of Research, Open Data Institute
Barbara **Ubaldi**, Head of Digital Government and Data Unit, OECD.

Chair: Stefaan **Verhulst**, The Gov Lab, New York University.

Panel Abstract

Our modern era is one of complex and interconnected problems, one where access to data has never been more crucial. Whether shaping public policy, responding to disasters, or empowering research, data plays a pivotal role in our understanding of complex social, environmental, and economic issues. Yet, progress on opening datasets has seemingly stagnated across much of the world, with fewer open data policies enacted and signed. In some cases, there has been backsliding; new limits on social media data for research, efforts to prevent the use of databases for generative AI training, and the privatization of climate data could be the start of a worrying trend that constrains open data and data reuse for public interest. Unless we speed up our efforts to secure data access, it is possible that we might end up in a "data winter," where there is reduced funding and interest in opening up data, akin to the AI winters of years past. Data in and for the public interest could be "frozen" by a lack of resources.

Panel: Session 5A

Better Together? How Smart Data and Public Data Can Co-Exist and Thrive

sponsored by

the ESRC Smart Data Research UK and the Alan Turing Institute

Speakers

Joel **Suss**, Data Journalist, Financial Times
Martine **Wauben**, Head of Data for London, GLA
Anya **Skatova**, Senior Research Fellow, University of Bristol
Blair **Freebairn**, CEO, GEOLYTIX

Introducers:

Mark **Birkin**, Director, Urban Analytics Programme, The Alan Turing Institute

Chair: Rachel **Franklin,** Newcastle University and The Alan Turing Institute

Panel Abstract

The rise of big tech and the production of smart data (as a by-product of human interaction with digital platforms) has vastly expanded the potential to address pressing policy questions using state-of-the-art methods like artificial intelligence. At the same time, the statistical reliability and representativeness of such data is in question. Recent changes to regulatory frameworks and changes to the corporate structures of big tech companies have exposed the vulnerabilities of relying solely on such datasets. These uncertainties are unfolding within a wider context of government responsibility to produce scientifically sound and statistically reliable datasets for public good.

So how do policymakers, stakeholders, researchers and industry best navigate these challenges to ensure the data ecosystem thrives? How do they meet the needs of a diverse range of constituents? This panel seeks to explore these questions and to consider the wider societal implications of data availability and uptake.

Panel: Session 6D

Interwoven Realms: Data Governance as the Bedrock for AI Governance

Speakers

Frederieke **Schüür,** UNICEF
Silvana **Fumega**, Global Data Barometer
Marta **Poblet,** The Data Tank
Andrew **Dwyer**, Royal Holloway Research Portal

Chair: Stefaan **Verhulst**, The Gov Lab, New York University

Panel Abstract

This panel will explore the relationship between data governance and AI governance, emphasizing how responsible AI governance is reliant on robust data governance and data quality practices. (Initial blog HERE)

Possible Discussion Points:

1. The role of data governance in shaping responsible and fit-for-purpose AI systems.
2. Addressing AI governance challenges such as bias and safety, regulatory adherence, and risk management through effective data governance practices.
3. Building trust and obtaining social license for AI systems through transparent and inclusive data governance.
4. Leveraging data governance and quality assessment frameworks as a common foundation for governing various data-driven technologies beyond AI - including the emergent focus on neurotechnology.
5. Lessons learned from data governance for the implementation and standardization of AI governance practices, and the importance of functions like data stewards..
6. Opportunities and challenges in strengthening global efforts to enhance data governance and stewardship for supporting ethical and trustworthy AI development

Panel: Session 7A

Delivering AI Assurance as a Service, Key Stakeholder Responsibilities organised in collaboration with

Validate AI and Imperial Data Science Institute

Speakers

Charles **Kerrigan,** Partner at CMS Legal
Tirath **Virdee,** Data and AI Professional
Zeynep **Engin**, Chair & Director, Data for Policy CIC

Chair: Ed **Humpherson**, Director General Regulation, Office for Statistics Regulation (OSR), UK

Panel Abstract

AI Assurance as a Service (AIAS or AIAaaS) represents a transformative approach in the burgeoning AI industry, crucial for policymakers to consider. It ensures AI systems in critical sectors, like healthcare and finance, adhere to the highest standards of safety and ethics. The approach uniquely balances technical expertise with vital human oversight, addressing the complexities of AI while ensuring ethical integrity. AIAS not only aligns AI development with regulatory standards but also fosters trust in AI technologies among the public. Its implementation encourages responsible AI innovation, providing a framework that integrates ethical considerations without stifling technological advancement. This approach is instrumental for policymakers aiming to promote a balanced, ethical AI landscape, ensuring technology serves societal needs while maintaining innovation and competitiveness.

The UK government announced in December 2021 that AI assurance will become a significant economic activity in its own right, with the potential for the UK to be a global leader in a new multi-billion-pound industry. The panel will debate how this goal will be achieved, to grow an *ecosystem* to build expertise where AI assurance services can thrive, delivering safe, trusted, and responsible AI.

We will explore the viewpoint that a community driven approach with diverse perspectives will yield the most robust, trusted AI outcomes. We will also counter this with a competitive approach to produce assurance models and which of the two or indeed a hybrid is best. Panellists will share their opinions from a legal, ethical, and technical stance as well as reflect on trends in the market both nationally and globally.

Relevant References:

UK GOV-CDEI AI Assurance Ecosystem Overview - Dec 2021
UK GOV - CDEI - Portfolio of AI Assurance Case Studies
Validate AI CIC - Our position on tackling AI Risks

Panel: Session 8A

Piercing the Veil: Technology's Role in Detecting Illegal Content

Speakers

Shubham **Jain,** PhD - Researcher at Imperial College London on online safety technologies and privacy
Andreas **Gutmann,** PhD - Senior Technologist Online Safety & Security at Ofcom
Rachel **Warner** - Barrister and former NCA investigator

Chair: Yves-Alexandre de **Montjoye,** Computational Privacy Group, Imperial College London

Panel Abstract

Lawmakers in the UK and EU have been pushing for technologies to be deployed to detect and remove illegal content such as Child Sexual Abuse Material (CSAM). The UK Online Safety Bill, which recently became law, mandates OFCOM, the national regulator, to detect and report the sharing of CSAM content to the police, including images shared in private end-to-end encrypted exchanges (E2EE). The EU CSAR bill, currently in trilogue, would require similar 'client-side scanning' technologies to be deployed.

The ability of client-side scanning technologies, which are based on a technique known as 'perceptual hashing', to safely and reliably detect illegal content in the E2EE context, has been called into question by recent research. Papers from Imperial College London and Georgetown U. have, for instance, demonstrated that client-side scanning technologies, whilst invasive, can be evaded by users with a degree of technical competence. A different study also identified how client-side scanning algorithms can be designed to be undetectably dual-purpose, for example hiding a facial recognition function as part of their image-scanning software.

Advocates argue that these tools are necessary to identify illegal images and those who share them, and that client-side scanning complements the reporting mechanisms that currently exist in non E2EE contexts. Others however argue that client-side scanning will undermine both the principle and practice of encryption, offering no guarantee that the dilution of online privacy for all users can be traded-off against the reliable protection of children online or the accurate identification of individuals sharing illegal images.

The goal of this panel would be to discuss in which context technology can or cannot help reliably and safely detect illegal content. This is a vital question for the regulators who are now being tasked with identifying the appropriate technologies to deploy as part of their new online safety responsibilities.

Panel: Session 9A

Historical Arc from WWII to AI: Collective Action for Global Equity

Speakers

Jude **Kong**, Assistant Professor, Department of Mathematics & Statistics, York University, Canada
Maximilian **Kasy**, Professor of Economics, University of Oxford *(Remote Participation)*
Rachel **Adams**, Global Center on AI Governance
Sanaz **Talaifar,** Imperial College London

Chair: Leigh **Anderson**, Marc Lindenberg Professor of Humanitarian Relief, International Development and Global Citizenship, University of Washington.

Panel Abstract

By some measures, global inequality has risen to post-WWII record levels, hypothesized to, among other consequences, be driving the rise of autocrats. (cite/#s) Amongst the optimistic narratives around "AI for Social Good", a pivotal question arises: how can we ensure AI is not used to concentrate power and further concentrate wealth?

Every new technology can be evaluated first in terms of how its use directly affects the quantity and sectoral distribution of other factors of production – land, labor, minerals - either to produce the same level of output at lower cost, or to increase output at comparable cost. As humans, labor is obviously different. Does technology's use increase the demand for labor (e.g. the steam engine), or does it replace labor (automatic teller machines), though perhaps creating a demand for new types of labor (maintaining or fixing ATMs). As a particularly shameful example, with the invention of the cotton gin, white settlers in the U.S. needed more land, so they stole it from native Americans, and they needed more labor, so they took it by force from Africa (quote from 1619). Second, how are the returns to that output or economic gains distributed between free labor and the owners of capital. Does the profit from cotton go to mill owners, or in the shameful past, owners of slaves? Or is the profit from cost-saving ATMs returned to shareholders? The answer, historically, is that returns have been skewed towards owners of capital, who are far more concentrated than the labor pool, hence rising inequality. (cite Branko Milanovic).

What is different between physical technologies – the printed book, steam engine, and cotton gin, and new digital technologies? One difference is the speed of diffusion and the high costs of containment. These attributes are also what distinguish global movements of capital from migration. And unlike trading in physical goods like cars, information also has the attribute of repeat use – something economists call nonrival consumption These attributes, particularly containment, create collective action problems.

AI and digital information technologies have another unique feature, namely the ability to change how individuals behave. The returns to controlling these technologies are therefore not only financial, they return power and control. The added dimension of AI over traditional social media and internet information sources, is the unrepresentative and biased corpus of information the algorithms pull from, and ability to prey on human cognitive tendencies towards confirmatory and other biases.

AI presents a monumental new collective action challenge on a global scale, yet our track record on similar challenges, such as climate change and international migration, leaves much to be desired. These distributional tensions have long been recognized, alternately resolved through conflict or cooperation. At points in time, such as post-WWII Bretton Woods, nations have come together to establish organizations and norms of engagement to avoid violence. But these agreements are only at a point in time, and with rogue defectors, we are once again witnessing violence as a means to concentrate power and wealth, made increasingly possible by the incremental increases in within nation inequality that erode young democracies in Africa and established ones elsewhere.

This session explores how we can surpass historical shortcomings in collective action since WWII, probing the effectiveness of systems like the UN that emerged from the Bretton Woods Agreement of 1945. Are these structures equipped to address the complex challenges we are facing around AI, or is there a need for novel approaches beyond their confines?

Panel: Session 9B

The Future of Data Ownership and Sovereignty:

An examination on current governance modalities and debate on anticipatory trends

(Area 3 - Designed Panel)

Speakers
Fei **Liao**, Nanjing Audit University *(Remote Participation)*
Yaniv **Benhamou**, University of Geneva,
Masaru **Yarime,** The Hong Kong University of Science and Technology
Ashraf **Shaharudin**, Department of Urbanism, Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, the Netherlands
Annelieke **van den Berg**, TNO (Netherlands Organisation for Applied Scientific Research)

Chair: Johanna **Walker,** King's College London

Panel Abstract
The Future of data ownership and sovereignty: An examination on current governance modalities and debate on anticipatory trends

The recently released "The Data for Policy Community" report (Engin et al. 2024) outlines the vital role of this distinct field and articulates the increased need to guide research under a proposed collective framework. It solidifies the research and stakeholder community around this interdisciplinary area on "Data for Policy", and expounds on 6 distinct yet interconnected themes to converge future research efforts and learnings. From this, the area on "Policy and Literacy for Data" focuses on the "the policy, governance and management issues involved in development and implementation of data-driven solutions" and the "governance models and frameworks for data and associated technologies [] developed across the globe with variances according to local context and value judgements, as well as public literacy and acceptance".

With this in mind, this panel discussion offers a brief reflection into current models, and an opportunity to learn how local contexts evolve with the changing data governance concepts with Latin America and China as specific examples, and emerging norms driven by rapid advancement in digital technology and the increased ambition from countries to provide a transformative approach towards a societal-scale digital public infrastructure to foster services and economic activites. It additionally reflects on anticipatory trends to come as data governance as a discipline and regulatory practice continues to change and

adopt to mega technological trends brought on suppliers and firms, and norms and principles debuted by countries and the international policymaking community.

This panel discussion will be opened and framed within a comprehensive effort of a landscape report underway within the context of the "Policy and Literacy for Data" area by its area committee to attempt a proposition in reflecting on said current dynamics, and proposing further questions and needs for understanding change in governance approach as technology, norms, individual and institutional literary shift with time. This panel will serve to complement and inspire the "Data governance, law and management of data" subchapter of said report, and the knowledge shared during the discussion and potential collaboration on the report with this set of invited authors will potentially supplement this directional-setting task to inform future questions.

References:

Engin Z., Gardner E., Hyde A., Verhulst S.V., Crowcroft J. (2024). Unleashing Collective Intelligence for Public Decision Making: The Data for Policy Community. Data & Policy. 2024;6:e2. https://doi.org/10.1017/dap.2024.2

Exploring the Contributions of Open Data Intermediaries for a Sustainable Open Data Ecosystem.

Ashraf **Shaharudin**,[1] Bastiaan van **Loenen**[1] and Marijn **Janssen**[2]

1. Department of Urbanism, Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, the Netherlands.
2. Department of Engineering, Systems and Services, Faculty of Technology, Policy and Management, Delft University of Technology, Delft, the Netherlands.

Sub. No: DAP-2023-0147
Full paper is accepted for publication in the Data & Policy journal.

Abstract

Open data promises various benefits, including stimulating innovation, improving transparency and public decision-making, and enhancing the reproducibility of scientific research. Nevertheless, numerous studies have highlighted myriad challenges related to preparing, disseminating, processing, and reusing open data, with newer studies revealing similar issues to those identified a decade prior. Several researchers have proposed the open data ecosystem (ODE) as a lens for studying and devising interventions to address these issues. Since actors in the ecosystem are individually and collectively impacted by the sustainability of the ecosystem, all have a role in tackling the challenges in the ODE. This paper asks what the contributions of open data intermediaries may be in addressing these challenges. Open data intermediaries are third-party actors providing specialized resources and capabilities to (i) enhance the supply, flow, and/or use of open data and/or (ii) strengthen the relationships among various open data stakeholders. They are critical in ensuring the flow of resources within the ODE. Through semistructured interviews and a validation exercise in the European Union context, this study explores the potential contribution of open data intermediaries and the specific ODE challenges they may address. This study identified 20 potential contributions, addressing 27 challenges. The findings of this study pave the way for further inquiry into the internal incentives (viable business models) and external incentives (policies and regulations) to direct the contributions of open data intermediaries toward addressing challenges in the ODE.

Capabilities for governmental data ecosystems for solving societal challenges.

Annelieke van **den Berg**[1] , Marissa **Hoekstra**[1]and Anne Fleur **van Veenstra** [1]
1. *TNO (*Netherlands Organisation for Applied Scientific Research)

Sub. No: 1320

Abstract

Introduction

Societal challenges such as climate change and public health require complex policy decisions. Governmental organizations increasingly have to rely on a good information position in order to cope with this complexity. Having access to data from various domains is seen as a facilitator for making evidence-informed decisions that are more legitimate and less uncertain. Often, a network of actors has to participate in order to identify and make data available that is stored at various organizations. These networks are also referred to as data ecosystems (Oliveira et. al, 2018). Various factors make it challenging to establish successful data ecosystems. Technical barriers are for example limited availability of data, poor quality of data and interoperability (Dawes, Vidiasova & Parkhimovich, 2016; Oliveira, Barros Lima & Loscio, 2019). Additionally, collaborating in networks asks for organizational and cultural capabilities. To unravel the main conditions that lead to effectiveness in this regard, we answer the question: What aspects contribute to the successful functioning of intergovernmental data ecosystems aimed at tackling societal challenges? Focusing on data ecosystems that are organized around societal challenges helps unraveling the complexity of creating a cultural shift towards working with data in collaboration and the surrounding governance processes.

Research methodology

A qualitative research design using case studies was deemed suitable to identify success factors of intergovernmental data ecosystems. Cases were selected for this study on the basis that the data ecosystem (1) involves participants from various governmental organizations, (2) is aimed at solving a societal challenge and (3) has a relatively high level of maturity. We included two cases from different domains, namely sustainability and safety. The selected cases were analyzed via a combination of desk research and semi-structured interviews with participants in the data ecosystem. For the desk research we analyzed program reports and other available documentation about the data ecosystems. For the interviews, we spoke with three participants per case, that were each employed within a different (governmental) organization and had different roles, such as program manager, director of information provision, researcher, and policy advisor.

Key findings

When data ecosystems are created with the goal of contributing to solving societal challenges it quickly becomes evident that there is a broad field of stakeholders to collaborate with, because they are either involved with the issue, or they hold data that are relevant to generating the necessary insights. Thus, the innovation that is necessary in data ecosystems is mostly focused on shaping this collaboration.

We find that the success of the data ecosystem is to a large part dependent on the extent to which parties in the data ecosystem are able to find common interests, decide which actions get prioritized, and figure out a way to work together. As one of the interviewees noted, the key challenge of sharing data within an ecosystem does not have so much to do with the technical aspects, but "it's about all the fuss surrounding it". In this section we untangle this fuss and share five narratives that describe how it can be successfully navigated.

1. Use the question as guide to articulate data demand

Both cases experienced that it is important to have a central demand or research question at the heart of the data ecosystem, that is derived from the societal issue. This research question can serve as the glue that holds the data ecosystem together and be guiding in identifying the next steps that need to be taken. Working with this approach requires a reversal from a supply driven to a demand driven mindset.

2. Invest in a common language and shared definitions

The complexity of data projects and societal challenges requires that policy and domain expertise, data expertise and legal expertise - across several institutions - come together in the initiative. This indispensable collaboration between multidisciplinary actors in a data ecosystem makes it a necessity to invest time and energy in making sure that all participants understand each other well. It is recommended to make room to ask and answer detailed questions, because otherwise there is a chance that information is produced everyone interprets in their own way.

3. Secure capacity through managerial support and a convener

Managerial support is highlighted as a crucial factor that impacts whether sufficient capacity is available to bring together all the necessary expertise in a data ecosystem. Data ecosystems surrounding societal challenges are vulnerable to fluctuations managerial support, because whether societal issues remain high on administrative agendas is dependent on various factors.

4. Foster collaboration through shared interests and expectation management

Flexibility of the participants is necessary, because at times the interest of the collaboration needs to be prioritized over the interest of the participants' institutions. The urgency of the societal challenge can allow participants to take on a more flexible attitude than they might otherwise. Additionally, making explicit what people's expectations are of participating in a data-sharing program and what their interest is helps to create rest and mutual respect.

5. Organically create structured ways of working

The organizational structure should emerge in an organic and bottom-up way that appeals to all participants. A convener role is necessary to match and coordinate knowledge and expertise between the different participating organizations.

These five narratives highlight the importance of 'softer' capabilities. Without these it is much harder to tackle the technical issues that may arise in a data ecosystem and create a prolonged and fruitful collaboration.

References

Dawes, S. & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. Government Information Quarterly, 33(1), pp. 15-27.

Oliveira, M., Oliveira, L., Batista, M., Loscio, B. (2018). Towards a meta-model for data ecosystems. dg.o '18: Proceedings of the 19th Annual International Conference on Digital Government Research. doi.org/10.1145/3209281.3209333.

Oliveira, S., Barros Lima, M.L., & Farias Lóscio, B. (2019). Investigations into Data Ecosystems: a systematic mapping study. Knowledge Information Systems 61, 589–630. https://doi.org/10.1007/s10115-018-1323-6.

Open Data Commons Licenses and Collective Data Governance for Personal and Non Personal Data.

Yaniv **Benhamou**[1] and Melanie **Dulong de Rosnay**[2]

1. *University of Geneva*
2. *CNRS (Centre national de la recherche scientifique)*

Sub. No: 4779

Abstract

This proposal following an article submitted to another journal includes policy recommendations and a standard template for open licenses applicable to any kind of data.Data are often subject to a multitude of rights (e.g. original works or personal data posted on social media, or collected through captcha, subject to copyright, and database and data protection) and voluntary shared through non standardized, non interoperable contractual terms. This leads to fragmented legal regimes and has become an even major challenge in the AI-era, for example when online platforms set their own Terms of Services, in business-to-consumer (B2C) relationship.

This proposal develops standard terms that may apply to all kind of data (including personal and mixed datasets subject to different legal regimes) based on the open data philosophy initially developed for Free, Libre and Open Source software and Creative Commons licenses for artistic and other copyrighted works.

Our work analyses how to extend open standard terms to all kinds of data. Wee suggest to combine these open standard terms with collective governance instruments, in particular data trust, inspired by commons-based projects and by the centennial collective management of copyright. Finally, we propose a template for Open Data Commons Licenses (ODCL), combining compulsory and optional elements to be selected by licensors, illustrated by pictograms and icons using legal design techniques and inspired by the bricks of Creative Commons licences.

This proposal aims at addressing the bargaining power imbalance and information asymmetry (by offering the licensor the ability to decide the terms), and conceptualises contract law differently. It reverses the current logic of contract: instead of letting companies (licensees) impose their own ToS to the users (licensors, being the copyright owner, data subject, data producer), licensors will reclaim the ability to set their own terms for access and use of data, by selecting standard terms. This should also allow the management of complex datasets, increase data sharing, and improve trust and control over the data. Like previous open licencing standards, the model is expected to lower the transaction costs by reducing the need to develop and read new complicated contractual terms. Last, it could spread the virality of open data to all data in an AI-era, if any input data under such terms used for AI training purposes propagates its conditions to all aggregated

and output data. In other words, any data distributed under our ODCL template will turn all outcome into more or less open data and build smaller or larger data common ecosystems.

Exploring Emerging Trends in Data Governance: An AI-Assisted Approach to Bibliometric and Text Analyses.

Mushan **Jin**[1] and Masaru **Yarime**[1]

1. *The Hong Kong University of Science and Technology*

Sub. No: 8454

Abstract

With the unprecedented speed and scale of data generation, storage, sharing, and reuse, data governance has been extensively discussed in both the intra-organizational context as well as from an inter-organizational and cross-functional perspective in the past two decades(Abraham et al., 2019; Zhang et al., 2022). However, data governance, which grew out of corporate governance, is still in its infancy and has not established a mature field of study(Jagals et al., 2021). The definition and framing of data governance remain ambiguous given the multidisciplinary nature of data and the inherently complex process of governance, and the term has yet to be clearly differentiated from synonyms such as data management or information governance(Jagals et al., 2021). Besides, the rapid changes in the social, technological, and environmental landscape complicate the conceptualization and practices of data governance models (Micheli et al., 2020). There is a need to recognize and understand the emerging concerns and challenges associated with such intersections, which require a comprehensive investigation of the evolution and direction of the field.

Aspects addressed in prior studies include data governance activities in both scientific and practice-oriented literature (Alhassan et al., 2016, 2018), data governance principles (Brous et al., 2016), data governance taxonomy (Al-Ruithe et al., 2018; Lis & Otto, 2021), data governance motivations(Walsh et al., 2022), data governance frameworks(Alsaad, 2023), cloud data governance (Al-Ruithe et al., 2019). A dynamic perspective of the developing body of knowledge on data governance is currently lacking (Abraham et al., 2019), despite nascent efforts to categorize insights through text mining techniques(Bozkurt et al., 2022). This study thus aims to provide a comprehensive and temporal overview of emerging trends and topics in data governance, contributing to a better understanding of its intellectual and conceptual structure. To achieve this goal, this research deployed an AI-assisted approach to bibliometric and text analyses to generate a holistic perspective on the delineation of research topics and emerging trends. This approach helps us avoid narrow searches and streamline the labor-intensive process of screening the massive literature. As part of our search strategy and screening stage, we applied semi-automated strategies to select input documents for our bibliometric analysis and text analysis. Bibliometric analysis was conducted to provide insights into the foundational studies and evolving trajectories of data governance research. Transformer-based pre-trained models were used to categorize the

input documents into different topics and analyze the associated changes in relevant topics over time. We selected the Web of Science and Scopus databases for our analysis, given their comprehensive coverage of academic literature and the rich bibliometric metadata provided. Data governance publications from 2004 to 2023 were covered.

Our results show that data governance research has increased exponentially over the past five years, with the main topics focusing on data standards, data architectures and organization, big data governance, open data, privacy and security, blockchain and consensus mechanisms, as well as advanced technology such as artificial intelligence (AI) and federated learning. In the top cited publications, frameworks or ethical principles are highlighted for diagnosing or designing data governance strategies, and privacy, data access, ownership, and stewardship of data are also discussed. Upon analyzing the trends in data governance topics, we found that some terms derived from corporate governance or information governance, such as data quality management and master data management, appeared relatively early and remained fairly constant throughout the study period. Comparatively, topics that relate to data ownership, regulations, privacy, and protection have increased rapidly since 2018, outpacing the earlier dominant technical and management-oriented topics. We also observed a shift in focus from digital identity regulation, copyright laws, localization of data, and personal data legislation to more custom data protection legislation in recent years, as indicated by the repeated occurrence of terms such as General Data Protection Regulation (GDPR). Data standards and governance topics related to health have increased significantly over the last three years as a result of the pandemic. Interestingly, the health-related data governance topics include terms related to monitoring and surveillance in addition to some common topics such as data sharing, data actionability, and dashboard construction.

The contributions of this study are twofold. First, we combined semi-automated bibliometric searches with a machine learning approach, which helps reduce the subjectivity of researchers and enhance search comprehensiveness. In the past, authors trying to grasp an overview of data governance usually used a keyword-based search and manual reference snowballing to select relevant articles, potentially causing a narrow search in the field of data governance and always time-consuming. Second, this study provides detailed insights into the changes in different topics in data governance from a temporal perspective through integrating bibliometric and text analysis techniques. Our findings suggest a growing interest in decentralized data governance, open data, privacy-preserving technologies, and data privacy regulations. The need for context-specific, decentralized, and collaborative models is also emerging in the existing discourse to reconcile the interests of different stakeholders (e.g., individual rights versus public interests). A taxonomy or conceptual framework of data governance can be further developed by balancing technical, ethical, legal, and social dimensions based on the entities identified in this study.

References:

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information Management, 49, 424-438.

https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2019.07.008
Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2018). Data governance taxonomy: Cloud versus non-cloud. Sustainability, 10(1), 95.
Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2019). A systematic literature review of data governance and cloud data governance. Personal and Ubiquitous Computing, 23(5), 839-859. https://doi.org/10.1007/s00779-017-1104-3
Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: an analysis of the literature. Journal of Decision Systems, 25(sup1), 64-75. https://doi.org/10.1080/12460125.2016.1187397
Alhassan, I., Sammon, D., & Daly, M. (2018). Data governance activities: a comparison between scientific and practice-oriented literature. Journal of Enterprise Information Management, 31(2), 300-316. https://doi.org/10.1108/JEIM-01-2017-0007
Alsaad, A. (2023, 14-16 Aug. 2023). Governmental Data Governance Frameworks: A Systematic Literature Review. 2023 International Conference on Computing, Electronics & Communications Engineering (iCCECE),
Bozkurt, Y., Rossmann, A., & Pervez, Z. (2022). A literature review of data governance and its applicability to smart cities. Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS 2022), 4-7 January 2022, virtual event/Maui,
Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016, 2016//). Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles. ELECTRONIC GOVERNMENT, Cham.
Jagals, M., Karger, E., & Ahlemann, F. (2021). Already grown-up or still in puberty? A bibliometric review of 16 years of data governance research. Corporate Ownership and Control, 19(1), 105-120. https://doi.org/10.22495/cocv19i1art9
Lis, D., & Otto, B. (2021). Towards a Taxonomy of Ecosystem Data Governance. https://doi.org/10.24251/HICSS.2021.733
Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. Big Data & Society, 7(2), 2053951720948087. https://doi.org/10.1177/2053951720948087
Walsh, M. J., McAvoy, J., & Sammon, D. (2022). Grounding data governance motivations: a review of the literature. Journal of Decision Systems, 31(sup1), 282-298. https://doi.org/10.1080/12460125.2022.2073637
Zhang, Q., Sun, X., & Zhang, M. (2022). Data Matters: A Strategic Action Framework for Data Governance. Information & Management, 59(4), 103642. https://doi.org/https://doi.org/10.1016/j.im.2022.103642

The National Audit as a Tool of Governance by Data' in China: A Phenomenological Approach.

Fei **Liao**[1], Mengjia **Gu**[1], Yi **Lu** [1] and Shichao **Zhou**[1]
1. *Nanjing Audit University, China*.

Sub. No: 6089 (Remote Presentation)

Abstract

Data generation, collection, storage, flow, analytics, and use are the fundamental elements of modern society. The process of datafication and algorithmization has penetrated into every facet of public administration (Broomfield & Reutter, 2022; Meijer et al., 2021; Vogl et al., 2020). Scholars have recognized the shift toward governance by data as a distinctive feature of governance in the digital age, and the associated transformation of political technology is emerging (Johns, 2021). The data and algorithmic ecosystems supporting decision-making in the public sector are the foundation of responsible governance. Therefore, we need to know more about "how to hold the data and algorithmic ecosystems accountable (Busuioc, 2020)," which is an essential but still overlooked issue. Establishing an independent third-party audit system is indispensable to address the challenges effectively.

Unlike prior literature, which mainly focuses on internal audit and contractual oversight, this article examines the governance by data, also the governance of data and algorithms from the novel lens of the national audit. Based on the phenomenological research of audit supervision practice in China, this article explores the role that national audit plays (or will play) as a tool of "governance by data", which is under steering of the Central Audit Committee of the Communist Party of China (CPC). First, when using data to fulfill their supervisory responsibility on public sector organizations, the national audit itself is a tool of "governance by data". Also, it is a high-level data user that identifies problems (errors, weaknesses, defects, and mistakes) in data and algorithmic systems of public sector and promotes improvements. Second, national audit supervises the performance of public funds invested in data and algorithmic systems and extends to public data assets themselves. Third, the leading cadres are responsible for implementing the digital transformation policy (including the policy of governance by data), and national audit works independently, objectively, and fairly to hold the cadres accountable for data and algorithmic ecosystems in the public sector.

This article also discusses the limitations of national audit on how to play its role in governing by data, including legal constraints and insufficient capabilities. The digital transformation of the national audit is a vital part of the overall technological shift in China's governance, which is also discussed in this paper. This original paper contributes to understanding how the CPC oversees data-driven governance initiatives.

Reference

Broomfield, H., & Reutter, L. (2022). In Search of the Citizen in the Datafication of Public Administration. Big Data & Society, 9(1). https://doi.org/10.1177/20539517221089302

Busuioc, M. (2020). Accountable Artificial Intelligence: Holding Algorithms to Account. Public Administration Review, 81(5), 825–836. https://doi.org/10.1111/puar.13293

Johns, F. (2021). Governance by Data. Annual Review of Law and Social Science, 17(1), 53-71. https://doi.org/10.1146/annurev-lawsocsci-120920-085138

Meijer, A., Lorenz, L., & Wessels, M. (2021). Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems. Public Administration Review, 81(5), 837-846. https://doi.org/10.1111/puar.13391

Vogl, T. M., Seidelin, C., Ganesh, B., & Bright, J. (2020). Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities. Public Administration Review, 80(6), 946–961. https://doi.org/10.1111/puar.13286

Panel: Session 9C

Designing a Value-driven GAI Framework for Social Good: Embedding Social Good Values into GAI Models

Speaker
Victor **Li,** The University of Hong Kong

Chair:
Jacqueline **Lam**, The University of Hong Kong

Panel Abstract
The increasing pervasiveness of Generative Artificial Intelligence (GAI) systems necessitates a re-examination of their ethical and social implications. This paper investigates existing GAI models, highlighting the challenges and limitations in their value alignment, followed by an investigation of how a value-centric GAI approach is essential for the ethical and moral development of future GAIs. By uncovering the implicit value assumptions of existing GAI models, we demonstrate how values have been embedded. Meanwhile, using finetuning and data generation techniques, we demonstrate how new values can be embedded in GAI models. Our study aims to advance understandings in implicit value assumptions in GAI models, paving the way for more ethical, responsible, and inclusive GAI model development. Our value-sensitive GAI model framework proposes a value embedding methodology detailing different techniques of embedding a new set of socially desirable values for social good.

AI at the Bench: Legal and Ethical Challenges of Informing – or Misinforming – Judicial Decision-Making Through Generative AI

David Uriel **Socol de la Osa**[1] and Nydia **Remolina**[2]

1. *Hitotsubashi University, Hitotsubashi Institute for Advanced Study, Graduate School of Law, Tokyo, Japan*.
2. *Singapore Management University, Singapore; Fintech Track Lead, SMU Centre for AI and Data Governance, Singapore*.

Sub. No: DAP-2023-0148
Full paper is accepted for publication in the Data & Policy journal.

Abstract

In this paper, we provide a systematic review of existing AI regulations in Europe, the United States, and in Canada. We build on the qualitative analysis of 129 AI regulations (enacted and not enacted) to identify patterns in regulatory strategies and in AI transparency requirements. Based on the analysis of this sample, we suggest that there are three main regulatory strategies for AI: AI-focused overhauls of existing regulation, the introduction of novel AI regulation, and the omnibus approach. We argue that although these types emerge as distinct strategies, their boundaries are porous as the AI regulation landscape is rapidly evolving. We find that across our sample, AI transparency is effectively treated as a central mechanism for meaningful mitigation of potential AI harms. We therefore focus on AI transparency mandates in our analysis and identify six AI transparency patterns: human in the loop, assessments, audits, disclosures, inventories, and red teaming. We contend that this qualitative analysis of AI regulations and AI transparency patterns provides a much needed bridge between the policy discourse on AI, which is all too often bound up in very detailed legal discussions, and applied socio-technical research on AI fairness, accountability, and transparency.

Risks and Best Practices for Using Generative AI in Judicial Decisions

Yuya **Ishihara**[1] and Mihoko **Sumida**[2]

1. *Hitotsubashi University, Faculty of Law*
2. *Hitotsubashi Institute for Advanced Study*

Sub. No: 8339

Abstract

The use of generative AI in the language domain has recently been expanding in a variety of fields of industries. A wide variety of use cases have been studied, ranging from sentence and code generation to retrieval of knowledge information and use as a dialogue partner in brainstorming sessions. As for the use of generative AI by judges, some cases are reported that generative AI supported to write the decisions and the possibility of generating sentences based on fictitious precedents by so-called hallucination is being investigated.
However, technologies have emerged to control and overcome the hallucination problems of generative AI, such as Retrieval Augmented Generation (RAG). Under current such situations, we conducted an investigation on possible use cases of generative AI by Judges at the litigation proceedings, not limited to those for which proof-of-concept research is underway in both public and private sectors, but also that are discussed as possibilities for the future. After presenting such use cases and specifying implementation patterns, this study discusses technical issues such as hallucinations, biases, and opacity of training data, which is also argued in some regulation, challenges that arise when using AI in civil court proceedings, and the risk of violating legal norms. It is important to take into account that the use of AI by judges in the court is also subject to the norm to which they are subjected, such as prohibition on the use of private knowledge of Judges, and that any violation of it involves not only ethical issues but also the risk of infringing the constitutional rights of the parties. With the ongoing discussion about the regulations on generative AI in mind, we argued desirable implementation design and propose the governance design of public legal RAG to the applications of generative AI in judicial decision making as a result.

Digital & Data-driven Transformations in Governance

AI Product Cards: A framework for code-bound formal documentation cards in the public administration

Albana **Celepija** [1,2,] Alessio Palmero **Aprosio** [1], Bruno **Lepri** [1] and Raman **Kazhamiakin**[1]

1. *Fondazione Bruno Kessler, Trento, Italy*
2. *University of Trento, Trento, Italy*

Sub. No: DAP-2023-0165 - St1
Full paper is accepted for publication in the Data & Policy journal.

Abstract

Currently, Artificial Intelligence (AI) is integrated across various segments of the public sector, in a scattered and fragmented manner, aiming to enhance the quality of people's lives. While AI adoption has proven great impact, there are several aspects that hamper its utilization in the public administration. Therefore, a large set of initiatives are designed to play a pivotal role in promoting the adoption of reliable AI, including documentation as a key driver.

The AI community has been proactively recommending a variety of initiatives aimed at promoting the adoption of documentation practices. While currently proposed AI documentation artifacts play a crucial role in increasing the transparency and accountability of various facts about AI systems, we propose a code-bound formal documentation framework that aims to support the responsible deployment of AI-based solutions. Our proposed AI Product Cards aims to address the need to shift the focus from data and model being considered in isolation to the reuse of AI solutions as a whole. By introducing a formalized approach to describing adaptation and optimization techniques, we aim to enhance existing documentation alternatives, thereby enabling the easy customization of AI solutions to specific contexts. Furthermore, its utilization in the public administration aims to foster the rapid adoption of AI-based applications due to the open access to the common use cases in the public sector. We further showcase our proposal with a public sector-specific use case, such as legal text classification task, and demonstrate how the

AI Product Card enables its reuse through the interactions of the formal documentation specifications with the modular code references.

How to design AI for public value: A socio-technical approach

Viviana **Bastidas**[1], Kwadwo Oti-**Sarpong**[1], and Jennifer **Schooling**[1]
1. *University of Cambridge*

Sub. No: 5049 -St1

Abstract

Cities are complex socio-technical systems. City managers are required to create public value and deliver socially desirable outcomes for all stakeholders (Bastidas et al., 2023b). The ambition to create artificial intelligence (AI) systems in cities that solve society's problems, align with the city's strategic goals, and adhere to ethical standards requires a holistic view of such socio-technical systems to ensure effective planning, design, and implementation (Yigitcanlar et al., 2021). However, AI projects in cities often fail because they focus mainly on technology, neglecting the complex interdependencies between people, strategies, policies, regulations, processes, and physical infrastructure. This has resulted in AI-enabled solutions that present significant risks and challenges, such as potential bias and discrimination, privacy violations and citizen surveillance (United Nations, 2022). According to recent studies, the application of AI technologies in urban landscapes can lead to harmful biased outcomes, depending on the training data used and the level of human supervision and control embedded in the design process (Floridi et al., 2021).

AI-driven systems are frequently encountered in day-to-day life. Their focus lies primarily on the efficient handling, storage, exposition, and utilisation of data (Cabrera et al., 2023). Government agencies have started to adopt AI to solve significantly complex tasks in diverse domains such as policy making, healthcare, transport, social welfare, public safety, and education (Androutsopoulou et al., 2019; Thanasis et al., 2022). When designed responsibly, AI can improve urban decision-making and interventions, inform governance, and public service delivery (Zuiderwijk et al., 2021). Many reports present the potential of AI for the public sector including data-driven decision-making, cost savings, enhanced public safety, and personalised services (Androutsopoulou et al., 2019). Despite these early applications and potential benefits, various ethical and social concerns impede the adoption of AI technologies, amplified by the increasing risks and the growing number of incidents (e.g. algorithmic errors, discriminatory outcomes, lack of transparency, privacy breaches, and cybersecurity threats) (AIAAIC, 2019). Various AI ethical regulations and principles have been introduced to address concerns related to bias, transparency, accountability, fairness, data privacy, and the ethical use of AI technologies across public and private sectors. However, the governance (planning,

management, operation, and use) of AI systems needs to move beyond high-level ethics guidelines and progress toward a more trustworthy process.

A strategy combining 'social' (urban governance and ethics) and 'technical' (digital technology) can enable designers of AI solutions to plan the development of AI-enabled systems that deliver social benefits while assessing and mitigating potential harms. This paper investigates the intrinsic relationship between the 'social' and 'technical' aspects of designing AI for public value. It critically reviews existing literature on urban AI-enabled systems and their associated digital architectures. This article draws on a socio-technical framework for responsible urban digital innovation (Bastidas et al., 2023a) that enables public value creation by encompassing three interrelated dimensions: governance and management, ethical and responsible innovation, and digital and technical. Examining the implications of AI systems design for the public sector holds significant relevance and urgency, due to the challenges faced by government agencies regarding the exploitation of AI capabilities and the management of potential risks (Fatima et al., 2022). Expanding the perspective beyond the 'technical' dimension of the design of AI systems, we investigate how the 'social' concept appears and is applied to designing AI systems and their digital architectures. This is fundamentally important to guide the responsible design and deployment of urban AI that enables public value creation.

The driving research question is: What are the socio-technical aspects of designing AI-based systems for public value creation? The main objective of this paper is to understand how these socio-technical concepts are used and applied to design AI systems and their digital architecture descriptions. It applies the concept-centric approach proposed by (Webster & Watson, 2002). In total 10 contributions were selected, representing a high-quality collection of journal and conference articles. Inspired by (Bastidas et al., 2023a; Bastidas & Schooling, 2024), we propose a conceptual 'socio-technical design of AI for public value' (see Figure 1) to determine the organizing and analytical framework of the review. The articles are reviewed according to their focus on different design components of AI systems. We define these design components as four distinct socio-technical layers: (1) the 'public value creation layer', (2) the 'governance and management layer' (3) the 'ethical and responsible innovation layer' and (4) the 'digital and technical layer'.

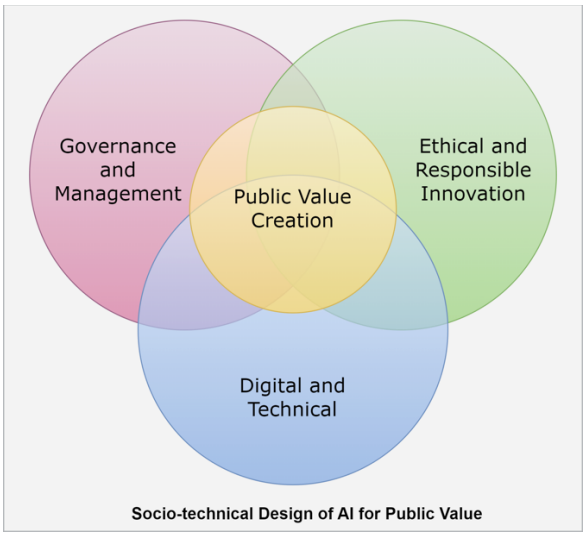

Figure 1. Socio-technical Design of AI for Public Value, Authors' Work.

The 'public value creation layer' describes the value-oriented features that AI-enabled systems should contribute to society (people and the environment). The 'governance and management layer' presents strategic and operational components needed to create public value enabled by AI technologies. The 'ethical and responsible innovation layer' constitutes the societal risk, impacts, and effects of implementing and deploying AI-enabled systems. The 'digital and technical layer' describes data, applications, and technology components of AI systems. Our review confirms the tendency in AI system design to prioritize the 'technical' while often neglecting the 'social' aspects of building it. We identify that while a few studies deal with data, algorithms, and public value propositions, they do not provide a deeper insight into the governance and management aspects that in turn affect public value creation. Some of these proposals provide only high-level ethical and responsible innovation elements for delivering fair and ethical systems. Most of the digital architectures are designed with the perspectives of 'technical' stakeholders in mind. However, the co-design of AI systems with 'social' stakeholders (e.g. citizens and communities) is missing. Furthermore, current studies have not yet developed interactions between the various components of our proposed socio-technical layers. This analysis underlines the need to adopt a socio-technical approach to designing AI systems for public value. This can ensure not only a common understanding of AI systems among all city stakeholders but also that the development of processes and tools that support the mitigation of potential harms can be implemented. A socio-technical approach to AI-based systems design can offer concrete recommendations for urban planners and developers on assessing risks and ensuring public value alignment.

References

AIAAIC. (2019, June). *AI, algorithmic and automation incident and controversy (AIAAIC) Repository*. https://www.aiaaic.org/aiaaic-repository

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, *36*(2), 358–367. https://doi.org/https://doi.org/10.1016/j.giq.2018.10.001

Bastidas, V., Oti-Sarpong, K., Nochta, T., Wan, L., Tang, J., & Schooling, J. (2023a). Leadership for responsible digital innovation in the built environment: A socio-technical review for re-establishing competencies. *Journal of Urban Management*, *12*(1), 57–73.

Bastidas, V., Oti-Sarpong, K., Nochta, T., Wan, L., Tang, J., & Schooling, J. (2023b). Leadership of urban digital innovation for public value: A competency framework. *IET Smart Cities*.

Bastidas, V., & Schooling, J. (2024). Socio-Technical AI Design for Public Value. *ECIS 2024 TREOS. 78.*, 1–3. https://aisel.aisnet.org/treos_ecis2024/78

Cabrera, C., Paleyes, A., Thodoroff, P., & Lawrence, N. D. (2023). Real-world Machine Learning Systems: A survey from a Data-Oriented Architecture Perspective. *ArXiv Preprint ArXiv:2302.04810*.

Fatima, S., Desouza, K. C., Buck, C., & Fielt, E. (2022). Public AI canvas for AI-enabled public value: A design science approach. *Government Information Quarterly*, *39*(4), 101722.

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2021). How to design AI for social good: seven essential factors. *Ethics, Governance, and Policies in Artificial Intelligence*, 125–151.

Popelka, S., Narvaez Zertuche, L., & Beroche, H. (n.d.). *Urban AI guide*.

Thanasis, P., Christou, I., Charalampos Ipektsidis, Soldatos, J., & Amicone, A. (2022). AI Solutions for Transparent, Explainable and Regulatory Compliant Public Policy Development Article title. *Data for Policy 2022*, 1–16.

United Nations. (2022). *AI and Cities Risks, Applications and Governance*. https://unhabitat.org/sites/default/files/2022/10/artificial_intelligence_and_cities_risks_pplications_and_governance.pdf

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, xiii–xxiii.

Yigitcanlar, T., Corchado, J. M., Mehmood, R., Li, R. Y. M., Mossberger, K., & Desouza, K. (2021). Responsible urban innovation with local government artificial intelligence (AI): A conceptual framework and research agenda. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(1), 71.

Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, *38*(3), 101577.

Investigating Public Sector Innovation Labs as-an-approach toward Data and AI-centric innovations in European National Governments.

Francesco **Leoni**[1], Stefano **Maffei**[1] and Bria Jammali-**Versace**[1]

1. *Department of Design, Politecnico di Milano*

Sub. No: 1039 -St1

Abstract

The need of a successful transition of public administrations towards the digital government ideal model increasingly compels public administrations worldwide to address challenges that stand far beyond the dimensions of organizational and technological innovation (Leoni et al., 2023). Earlier e-government studies emphasized the importance of intra-governmental integration of public information systems and of specific ICT-centered solutions (Charalabidis et al., 2019). Today, a more contemporary sensibility seems to reinforce the need to look outside governmental boundaries in the digital transformation of the public sector (Ravšelj et al., 2022).

In fact, whilst boasting of positive potential to transform the paradigm of public, digital transformation exhibits the markings of a socio-technical problem, i.e., a problem that speaks to all public issues of high impact that are void of potential for incisive problem identification and solving (Rittel & Webber, 1973).

Data-centric public services and AI-based solutions in the public sector are, therefore, increasingly addressed as socio-technical challenges that require broad-level considerations on data ethics, algorithmic legibility, social acceptance of technology, and coordination across public bodies. It is expected that new forms of collaboration between government and other societal actors will emerge based on organizational and semantic interoperability, thus suggesting the need to experiment with new forms of governance based on co-designed and participated processes with the ecosystem of stakeholders and beneficiaries (citizens).

However, the public sector is still characterized by a functional 'silos' model, with a fragmentation of competencies and mandates (in Italy, for example, the National Statistical Agency counted more than 12,800 public bodies in a 2017 census). Several analyses carried out by international observatories indicate that the adoption of digital and data-driven solutions can improve the productivity and resilience of the public sector, as well as the perceived quality of its services (Ubaldi et al., 2019) when accompanied by horizontal organizational integration based on new institutional formulas, coordination mechanisms and policy tools that support a whole-of-government approach to digital governance (Dener et al., 2021).

Europe is encouraging this perspective with a series of dedicated strategies, which foster collaborative governance among public actors, inclusive towards other social partners, especially towards citizens, the ultimate beneficiaries of the digital transformation in PA (e.g., Data Governance Act). In this sense, it is also worth mentioning The European Digital Rights and Principles and the EU 2030 Policy Programme, whose vision of digital transformation is functional to a transition to a climate-neutral, circular, and resilient economy, to be achieved by "[...] pursue digital policies that empower people and businesses to seize a human centred, sustainable and more prosperous digital future." (EC, 2021, p.1).

In response to these challenges, public sector innovation labs (PSI Labs) or policy labs have been introduced in many countries, whose purpose is to research and test innovative practices and approaches for the transformation of the public system (McGann et al., 2021). In recent years, this phenomenon has become increasingly widespread internationally with different models of action, often as organizational units (teams) within the public sector function with a specific mandate to experiment with new forms of innovation related to governance and services. There are now several concrete examples of PSI Labs in various national states, implemented both within national ministries and agencies (e.g., the Laboratorio de Gobierno in Chile or LabX in Portugal) and in public and territorial agencies (e.g., the 27eme Région in France). In this sense, rather than identifying an absolute typology, PSI Labs seems to obey organizational constraints and opportunities peculiar and contextual to the ecosystems of subjects and practices in which they are introduced (Lindquist & Buttazzoni, 2021). Their establishment should, therefore be understood starting from a precise relation with a given institutional/public context.

On these premises, this paper proposes a study of PSI Labs as-an-approach; in other words, PSI Labs as an action of governmental bodies toward public sector digital transformation. While several mapping and listing of PSI Labs exist, little research that concentrates on how PSI labs can be used to address the complexities of digital transitions while affecting policymaking (Carstens, 2023; Kim et al., 2022; Sandoval-Almazan & Millán-Vargas, 2023)

To investigate this background, we ask the following: (RQ1) What are the main typologies of projects undertaken by PSI labs dealing with digital transformation at the central government level? (RQ2) What are the main characteristics of PSI lab as-an-approach to data/AI-centric innovations in the public sector? (RQ3) How are public bodies influencing policymaking through digital government initiatives by adopting PSI lab as-an-approach?

To answer these questions, we developed a qualitative analysis of desk research data regarding the project portfolio of 6 PSI Labs working within, or in close relation with, the central government (i.e., public agencies or in-line departments) across 6 different European countries (Germany, Portugal, Norway, United Kingdom, France, Scotland).

References

Charalabidis, Y., Loukis, E., Alexopoulos, C., & Lachana, Z. (2019). The Three Generations of Electronic Government: From Service Provision to Open Data and to Policy Analytics. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11685 LNCS(February 2020), 3–17. https://doi.org/10.1007/978-3-030-27325-5_1

Dener, C., Nii-Aponsah, H., Ghunney, L. E., & Johns, K. D. (2021). GovTech maturity index: The state of public sector digital transformation. World Bank Publications.

Leoni, F., Carraro, M., McAuliffe, E., & Maffei, S. (2023). Data-centric public services as potential source of policy knowledge. Can "design for policy" help? Transforming Government: People, Process and Policy, ahead-of-p(ahead-of-print). https://doi.org/10.1108/TG-06-2022-0088

McGann, M., Wells, T., & Blomkamp, E. (2021). Innovation labs and co-production in public problem solving. Public Management Review, 23(2), 297–316. https://doi.org/10.1080/14719037.2019.1699946

Ravšelj, D., Umek, L., Todorovski, L., & Aristovnik, A. (2022). A Review of Digital Era Governance Research in the First Two Decades: A Bibliometric Study. Future Internet, 14(5), 126.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. Policy Sciences, 4, 155–169. https://doi.org/10.1080/01636609209550084

Ubaldi, B., Van Ooijen, C., & Welby, B. (2019). A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance. OECD Working Papers on Public Governance, 33, 1–59. https://doi.org/10.1787/09ab162c-en

Data Governance in Data Altruism: Archetypes Definition

Federico **Bartolomucci**[1], Edoardo **Ramalli**[1] and Valeria Maria **Urbano**[1]

1. *Politecnico di Milano, Italy.*

Sub. No: 1676 -St1

Abstract

The European Union, with the Data Governance Act, introduced a new status recognition, the Data Altruism Organization (DAO), that qualifies a not-for-profit and legally independent entity to offer data sharing, intermediation and analysis services for altruistic purposes. Their success in managing the partnership depends on several intertwined organizational and technological factors. However, the Data Governance Act provides high-level guidelines on DAOs' scope of work while leaving space for proposals and experimentation on potential governance configurations they may assume inside data ecosystems. Adopting a holistic perspective on data governance that integrates technological and organizational dimensions, our research put forward a conceptual framework identifying different DAO archetypes. For each archetype, we describe potential configurations as combinations of organizational and data governance aspects. Findings suggest the possibility for DAOs to adopt three alternative configurations depending on the role they assume for the collaborative ecosystem: the Facilitator, the Infrastructure Provisioner, and the Data Knowledge Centre. This work facilitates practitioners adopting data altruism by providing a set of possible data and organizational governance configurations. At the same time, the framework developed, which links and highlights the interdependence between the organizational and technological aspects of data governance, constitutes a valuable framework for analyzing other data-sharing contexts, adopting the same holistic approach.

Exploring the Intersection of Political Orientation and AI Governance Research: A Comprehensive Analysis of US Think-Tank Publications Using Large Language Models.

Emily Eunji **Kim**
Georgia Institute of Technology

Sub. No: 4761 -St1

Abstract

As artificial intelligence (AI) continues to reshape societal landscapes, understanding the influence of political orientation on AI governance research becomes increasingly critical. This study delves into the complex interplay between political ideologies and the discourse surrounding AI governance, focusing on publications from prominent US think-tanks. Leveraging advanced large language models, we analyze a vast corpus of texts to uncover patterns, biases, and divergences in perspectives.

Our methodology involves the application of cutting-edge natural language processing techniques to systematically examine and categorize publications from a diverse range of think-tanks. By harnessing the power of large language models, we aim to identify the implicit ideologies and underlying sentiments within the discourse surrounding AI governance. The analysis encompasses a spectrum of topics, including policy recommendations, ethical considerations, and regulatory frameworks.

Preliminary findings indicate a nuanced relationship between political orientation and AI governance stances within think-tank publications. The research sheds light on the potential impact of political ideologies on proposed AI policies, revealing alignments and disparities across the political spectrum. Understanding these dynamics is crucial for fostering informed decision-making and equitable governance in the era of AI.

This study contributes to the broader conversation on the societal implications of AI by providing empirical insights into the intersection of political ideologies and AI governance discourse. The findings have implications for policymakers, researchers, and stakeholders involved in shaping the future of AI governance, emphasizing the need for a comprehensive and inclusive approach that considers diverse perspectives across the political spectrum.

AI Documentation Method with Databook: Case Study of a Fraud Detection Model Audit.

Anna **Nesvijevskaia**[1] and Simon Le **Mouellic**[2]

1. *Conservatoire National des Arts et Métiers - DICEN Ile-de-France*
2. *Quinten*

Sub. No: 8094-St1

Abstract

The debates on the transparency of algorithms are in full swing within the European Union with the recent adoption of the Artificial Intelligence Act which follows the Digital Services Act and the General Data Protection Regulation. In addition to dissuasive fines for uses that are unacceptable and incompatible with the European values under discussion, such as social scoring or behavioural design of which citizens would be unaware, the documentation of algorithms appears as one of the main levers of transparency. The classic back-and-forth between innovation and regulation in this context is mainly aimed at foreign organisations, whose rapidly expanding practices (platforms, social networks, quantified-self movement, etc.) must be aligned with the values desired by European citizens. It also affects local companies that are gradually appropriating Artificial Intelligence algorithms. Indeed, stemming from technological and analytical evolutions propelled by a mythology (Boyd and Crawford, 2012), the spread of these techno-economic objects in compagnies has accelerated over the last dozen years, supported by the massive commercialisation of Cloud Data Lake technologies and Data Science projects. The first aim to reduce IT costs while exploiting a greater variety of data. The second aim to create business usages based on algorithms to generate productivity gains or new activatable knowledge. These projects are beginning to lead to concrete applications in the most mature, but often remain exploratory through *agile* methods that neglect documentation. Thus, the current context is marked by a twofold tension in companies: documentation practices are insufficient in the algorithm co-design phase to meet future regulatory requirements, and the usages developed in the past need rapid redocumentation for audit issues.

The tension in the design phase has been addressed in some French companies[3], where the field shows that the documentation of algorithms is a formidable tool for Human-Data Mediation during Data Science projects: it helps to facilitate coordination between actors with heterogeneous interests, skills and social universes (Arruabarrena et al., 2019) and generates a project memory for the capitalisation of business knowledge and

methodological skills for the management of data project portfolios (Nesvijevskaia, 2019). This observation led to the development of a boundary object for dynamic documentation of algorithms and associated metadata, called Databook, which has gradually evolved into a standard framework espousing the adjusted Data Science process model CRISP_DM (Chapman, 1999; Wirth and Hipp, 2000). It is structured as follows: each of the 6 activities of the CRIMP_DM model is associated to a specific data object on the critical path of the algorithm design (business concepts of the project perimeter, source data, model structure, analytical results, Functional results and usages, including direct actions and indirect knowledge to deploy); each data object is associated to metadata generated throughout the design process carried out by different stakeholders from a variety of angles, including statistical, semantic, technical, regulatory or business impact; this qualification leads to a methodical inclusion and exclusion of object data in the final algorithm. This simple, comprehensive metadata structure accelerates convergence between heterogeneous stakeholders during a Data Science project, supports the quality of the final algorithm and documents it at reduced cost, including in terms of arbitrations through the Databook versioning and excluded data objects. It also enables knowledge capitalisation based on the type of skills involved in the project and facilitates the re-use of data qualification across a portfolio of data projects. The Databook has been tested on hundreds of data projects in the design phase, approved and fully described and shared in actual Excel format (Nesvijevskaia, 2021).

In the recent context of imperative documentation of key usages involving AI, this framework, compatible with any project management method, offers an operational solution for creating transparent algorithms *by design* adhering to the FAIR principles (Hagstrom, 2014). However, this documentation method has limitations, such as improvable ergonomics or the need to redefine stakeholder roles for each project. These limitations are being addressed in practice through gradual adaptations, iteratively tested against scientific communities (Nesvijevskaia, 2023) and theoretical documentation approaches. A natural adaptation has been combining the Databook metadata model with practical documentation formats, such as text reports in Word or PowerPoint, with metadata included in figures and tables. For example, in 2022, a French bank used this approach while deploying a fraud detection AI model (Nesvijevskaia et al., 2021). The combined documentation was efficient for the design, deployment, and exploitation of the algorithm, but also facilitated the audit by the Data Protection Officer. This audit was carried out in 2023 through the application of the French Data Protection Authority (CNIL) guidelines for AI audit, consisting in 199 thematical questions. Documentation eases all these questions, with 50% relying on algorithm metadata covered by the Databook framework. The remaining questions can be limited to more declarative answers, perfectly covered by standard documentation formats. This combined documentation bridges the gap between the audit requirements and algorithm-specific documentation.

At this stage of our research, we compared this use case with audit and project management literature to develop a comprehensive AI documentation method. This method includes a shortlist of documents, which are collaboratively created artefacts (Gagnon-Arguin et al., 2015; Zacklad, 2013). These documents are linked to audit subjects necessary to ensure algorithms transparency and reduce associated risks and misuse. We articulate this

document shortlist with algorithm metadata documentation using the Databook framework. Beyond our first live case in fraud detection, we discuss this AI documentation method in terms of practical application, answers to major economic and social issues, and applicability for projects that aim to generate LLM models. Indeed, the critical path of data objects associated to flexible metadata should be suitable for these models, but the practical relevance must be tested for interactivity with AI model users, usage multiplication and evaluation difficulties. Finally, we discuss potential limitations and alternative audit methods (in-depth expert audit, labels and certifications, platforms for evaluating algorithms…), as algorithm documentation becomes urgent for European organisations and their partners.

Acknowledgments. The authors are grateful for the support provided by Quinten, for colleagues' reviews as well as for the long collaboration with our partner Banque Populaire Rives de Paris who shared expertise.

Funding statement. This research was supported by Quinten, exclusively covering the mobilised human resources. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests. Anna Nesvijevskaia and Simon Le Mouellic are employed at company Quinten.

Data availability statement. The data that support the findings of this study are available from Quinten's partner Banque Populaire Rives de Paris. For reasons of confidentiality, we cannot provide these data.

Author contributions.

- Anna Nesvijevskaia: Conceptualization, Methodology, Investigation, Validation, Writing - Original Draft, Writing - Review & Editing, Software, Data Curation, Formal analysis, Visualization, Supervision, Funding acquisition, Project administration.
- Simon Le Mouellic: Writing - Review & Editing, Software, Investigation, Data Curation, Formal analysis, Visualization.

References

Arruabarrena, B., Kembellec, G., Chartron, G., 2019. *Data littératie & SHS : développer des compétences pour l'analyse des données*, in: CODATA. Marne la Vallée, France.

Boyd, D., Crawford, K., 2012. Critical Questions for Big Data. *Information, Communication & Society* **15**, 662–679. https://doi.org/10.1080/1369118X.2012.678878

Chapman, P., 1999. *The CRISP-DM User Guide*.

Gagnon-Arguin, L., Mas, S., Maurel, D., 2015. *Les genres de documents dans les organisations : analyse théorique et pratique*, Gestion de l'information. Presses de l'Université du Québec, Québec, Canada.

Hagstrom, S., 2014. *The FAIR Data Principles* [WWW Document]. FORCE11. URL https://www.force11.org/group/fairgroup/fairprinciples (accessed 2.22.20).

Nesvijevskaia, A., 2023. Documentation of algorithms with Databook: from co-design to audit issues. DOCAM 2023 Annual Meeting, *Document Design and Co-Design*, CNAM, Paris.

Nesvijevskaia, A., 2021. DATABOOK : a standardised framework for dynamic documentation of algorithm design during Data Science projects. *IASSIST Quarterly,* **45**. https://doi.org/10.29173/iq989

Nesvijevskaia, A., 2019. *Phénomène Big Data en entreprise : processus projet, génération de valeur et Médiation Homme-Données* (thesis). Paris, CNAM.
Nesvijevskaia, A., Ouillade, S., Guilmin, P., Zucker, J.-D., 2021. The accuracy versus interpretability trade-off in fraud detection model. Cambridge University Press, *Data & Policy* **3**. https://doi.org/10.1017/dap.2021.3
Wirth, R., Hipp, J., 2000. CRISP-DM: Towards a standard process model for data mining, in: *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*. pp. 29–39.
Zacklad, M., 2013. Sémiotique de la création de valeur dans l'économie des transactions coopératives, in: Aggeri, F., Favereau, O., Hatchuel, A. (Eds.), *L'activité marchande sans le marché ?* : Colloque de Cerisy, Économie et gestion. Presses des Mines, Paris, pp. 265–283.

Catching the bad apples to keep up the good work: Dutch municipal government perspectives on data-driven governance.

Margot K**ersing**[1], Lieke **Oldenhof**[2], Kim **Putters**[3] and Liesbet van **Zoonen**[4]

1. *PhD Erasmus University Rotterdam*
2. *Associate Professor Erasmus University*
3. *Professor Tilburg University*
4. *Professor Erasmus University*

Sub. No: 5934-St1

Abstract

Despite the rapid rise of the digital welfare at the local government level in the Netherlands, there seems to be still surprisingly little political debate about the implications of digitalization for the life of citizens and the work of frontline bureaucrats. Only when there is an incident, such as the Dutch childcare benefits scandal, the SyRI social welfare fraud detection algorithm that violated the ECHR, or Rotterdam's discriminatory benefit fraud risk-assessment algorithm, the municipal government pays attention, but broader political debate on technology has yet to get off the ground (Rathenau Instituut, 2020, Open Rotterdam, 2023, Algorithm Watch, 2020, Henley, 2021). The current dualized structure of public administration at the municipal level in the Netherlands poses an obstacle to discussion about responsible use of data-driven technologies. Although the municipal council is tasked with setting the course, the implementation lies with the administrative body. In the Netherlands technology is viewed merely as an implementation issue and is therefore handled by aldermen and civil servants (Centre for BOLD Cities, 2023). According to the Association of Dutch Municipalities (VNG) it is necessary for the legitimacy of data projects in the public sector that local politicians are more involved in discussing the development of data-driven governance because in digitization processes public values like transparency, legitimacy, and privacy are under pressure (Verhoeven 2019 p.7, in VNG 2021, VNG, 2022). Furthermore, earlier research indicates that the use of data-driven technologies can negatively affect the work of frontline professionals (Kersing et al., 2022) and citizens' lives (Kersing et al., forthcoming). We argue that digitization of welfare provision is not merely a bureaucratic challenge but also a political one.

Current literature provides limited insight into local politicians' involvement in discussions about digitalization processes within the bureaucratic organization. Local politicians are rarely involved in the digitalization of service provision due to their lack of knowledge and oversight (VNG, 2021, Centre for BOLD Cities, 2023). Civil servants' reluctance to inform politicians often stems from a fear of politicians who want to avoid a negative image, and thus perhaps thwart the development of data projects (VNG, 2021).

In this research we explore municipal government discussions on the use of data-driven technologies used in the domain of work and income in the municipality of Rotterdam. Even though the municipality of Rotterdam is a frontrunner compared to smaller municipalities when it comes to experimenting with data-driven technologies, their use in the domain of work and income is relatively new and not without problems. In recent years the municipality has been criticized by the local audit office and investigative journalists for the negative consequences for citizens of their algorithm use (Rekenkamer Rotterdam, 2021, Open Rotterdam, 2023).

Therefore, we pose the following research question: How is the use of data-driven technologies in the domain of work and income politically discussed within the municipal government of the municipality of Rotterdam?

The aim of this study is to gain insight in if and how discussions about the use of data-driven technologies were politicized or not. We use the literature about (de)politicization strategies (a.o. Eliasoph and Lichterman, 2018) because these strategies enable us to indicate three important aspects of political discussions: the where, what, and how. Firstly, where discussions are taking place refers to the physical place such as political arenas, agencies, boards, and commissions. Secondly, they indicate what is discussed. By politicization and depoliticization actors influence what issues are up for deliberation and which ones are not. Thirdly, thematic types of depoliticization used by actors give an indication of how issues are discussed (in a political, economic, technological, ethical, or legal way).

To answer this question, we used a quantitative text-analysis software ConText and qualitative analysis software Atlas.ti to do a document analysis of municipal council and committee documents, and a content analysis of video recordings of council and committee meetings. We took an abductive, exploratory approach to make sure we would get a general view of how the municipal council discusses the use of data-driven technologies both before and after scandals occur over a period of 8 years (2016-2023). The document analysis includes documents and video recordings of the municipal council (2016-2023) and three municipal committees in the domain of work and income (2016-2018, 2018-2022, 2022-2023).

Our results show that data-driven technologies are used in the domain of work and income because their outcomes end up for example in the monitor work and income that is discussed every four months. However, there are rarely discussions or critical reflections in the municipal council on how the data-driven technologies are used.

On the rare occasions that they are discussed in the council it is typically in response to (1) scandals such as in the childcare benefits scandal and SyRI, or (2) criticism from for example audit offices. Most discussions were shifted towards- or exclusively discussed in committees.

Local politicians used depoliticization strategies to deal with criticisms such as (1) shifting issues from the political or ethical sphere to the scientific sphere (thematic framing), (2) diffusing responsibility across various actors, thereby distancing their own responsibility and blurring accountability.

Furthermore, we identified a strategy that we called 'content chopping', where issues are chopped into small content pieces (technical, ethical, political, executive) and spread into separate documents and discussion arenas. Thereby obscuring the overall coherence which diffuses critical concerns.

Another strategy we identified, the 'depersonalization strategy', involves abstracting issues to the point where human elements are lost. For example, in a technical session explaining the workings of a fraud risk prediction algorithm, characteristics of benefit recipients were replaced by characteristics of apples. An unfortunate comparison likening potential fraudulent benefit recipients to 'apples used to make applesauce' was used, implying that certain characteristic, lead to its unsuitability for sale and its processing into apple sauce. This metaphor, referring to the 'one bad apple can spoil the barrel' saying, suggests a negative view towards benefit recipients.

GRAIL: Developing responsible practices for AI and machine learning in research funding and evaluation with a community of learning.

Denis **Newman-Griffis**
University of Sheffield

Sub. No: 4190-St1

Abstract

Research/Policy Question
Research funding agencies are tasked with bridging the gap between the dual, rapidly-evolving areas of public policy and scientific advancement. As the growth of the scientific community continues to outpace uplifts in funding allocations, and increasingly data-driven policy creates demands for stronger evidence of efficacy and impact, there is a clear need for better tools to enable funders to ensure their decision-making stays well aligned with current science and policy needs.
The maturation of artificial intelligence (AI) and machine learning (ML) technologies is opening new avenues for funders to learn from the rich data sources and internal expertise they have curated over decades, and to develop new data-driven practices to support responsiveness to rapid scientific development as well as changing policy environments. However, there is a lack of shared experience and best practice in using AI and ML in the work of research funding and evaluation, and it is often unclear how developments in AI Safety and Responsible AI discourses translate into practical insights for complex organisations like research funders.
The Research on Research Institute's GRAIL project (Getting Responsible about AI and Machine Learning in Research Funding and Evaluation) is an ongoing effort drawing on a community of learning among research funding organisations to develop specific insights, pathways, and critical questions to guide responsible use of AI and ML in the research funding ecosystem. The GRAIL project brings together nine government research funders (Australian Research Council, Austrian Science Fund, Dutch Research Council, German Research Foundation, Research Council of Norway, Social Sciences and Humanities Research Council of Canada, Swedish Research Council, Swiss National Science Foundation, UKRI) and four philanthropic funders ("la Caixa" Foundation, Novo Nordisk Foundation, Volkswagen Foundation, Wellcome Trust), each of whom is at different points in the process of exploring, adopting, deploying, and evaluating AI and ML approaches in their work. The discussions and investigations in the project aim to produce new insights into how research funders and other public bodies can effectively navigate the sociotechnical systems and processes required to bring AI and ML technologies to bear effectively in their work whilst maintaining the highest standard of ethics and social responsibility.
Methodology
The GRAIL project consists primarily of a series of virtual, co-productive workshops held with staff from the participating organisations. Each workshop is hosted by one of the

project partners and is organised around a specific topic regarding an area of AI/ML application in research funding and/or a particular challenge in effectively and ethically managing AI/ML use.
Workshops are closed sessions with limited external data sharing and a strong focus on protected conversation with the freedom to discuss challenging topics and experiences. The host organisation for each workshop may invite external presenters and additional guests as relevant, with all attendees agreeing to abide by a co-produced set of ground rules. Workshop discussions are noted by the project team, with anonymised versions of notes produced for sharing to attendees after the workshop. Attendees are also invited to complete an anonymous feedback survey reflecting on the presentations/activities and discussion topics in each workshop and highlighting specific learning to carry forward. Workshop notes are reviewed by the project Steering Group, consisting of nominated representatives from participating organisations, and analysed by the Steering Group together with the project team. Emergent themes and recurring topics are identified by group discussion and used to update evolving documentation of key topics and challenges in navigating implementation and management of AI/ML in research funding and evaluation. As the workshop series continues, discussions will increasingly reflect specifically on this developing documentation and the insights and recommendations produced.

Key Findings

The current workshop series has developed to respond to and expand upon the insights generated from a three-session set of workshops in 2021, summarised in a recent RoRI report (Holm et al., 2022). Five GRAIL workshops have been held as of May 2024, on: (1) generative AI in the research funding ecosystem; (2) using AI in ex post evaluation of research outputs; (3) developing meaningful AI/ML guidance for funders; (4) NLP in research funding; and (5) policy and responsible AI. Further workshops are being scheduled for early 2024, with at least four sessions anticipated prior to the Data for Policy conference in July, on topics including (a) evaluating efficacy and change processes in introducing AI for reviewer matching; and (b) building resilient change processes to adapt to evolutions in AI technologies.
Key themes that have emerged from discussions to date include:
Theme 1: Disconnects between AI performance evaluation and organisational impact. It is not clear when an AI model that has been developed can be considered “good enough” to use, or how to effectively measure the balance of risk between AI adoption and continuing with current processes. The question of measuring reliability of AI systems, and developing organisational and sector understandings of what reliability should mean in the context of research funding and evaluation, is also a significant challenge that must be explored.
Theme 2: AI as one of many tools. Use of AI is often discussed as a direct improvement of (or replacement for) existing decision-making processes, however workshop attendees have highlighted getting more value from AI as a tool for process insight than efficiency improvement in many cases.
Theme 3: Managing across competencies. AI use is not a technical problem alone, but must combine experiences and perspectives from technical, strategic, and operational components of funding organisations. New work is needed to develop best practice for

interprofessional team management and communicating across competencies within organisations to achieve effective AI use.

The ongoing community of learning in the GRAIL project is well positioned to build on these initial themes to develop best practice and greater insight for using AI and machine learning in research policy ecosystems.

References

Holm, J., Waltman, L., Newman-Griffis, D. & Wilsdon, J. (2022). Good practice in the use of machine learning & AI by research funding organisations: insights from a workshop series. RoRI Working Paper No. 10. December 2022. DOI: 10.6084/m9.figshare.21710015

Data-driven analysis of school performance measurement

Ian **Widdows**
University of Sheffield

Sub. No: 9082-St1

Abstract

Research Questions

The evaluation and critique of the use of secondary school performance measures (SPMs) have been the subject of academic research throughout the three decades that they have formed part of the English educational system. They currently attract considerable attention within a wider debate around the suitability and effectiveness of the existing school accountability system. However, whilst increasing numbers of educational providers are collecting richer and richer data on education context, delivery and outcomes, these data rarely contribute to the use of SPMs in a policy context. In this study the use of empirical data and digital technologies contribute to an assessment of the extent to which SPMs are meeting their stated policy goals.

The research questions for this study are: How can we empirically conceptualise the underlying latent constructs shaping 'performance' within SPMs?

With reference to the current headline performance measures, what is the extent and breadth of the concept of 'performance'?

To what extent do SPMs encapsulate elements beyond the influence of schools and can digital tools aid in assessing their influence on the stated policy goals of SPMs?

The findings of this study will contribute to wider research which explores the effects and effectiveness of SPMs and seeks to develop new, more data-driven models for them. This will include the novel use of digital technologies, including AI, to synthesise more multidimensional data on school performance.

Research Methodology

The methods used in this work are exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and correlation analysis. In addition, this study is being extended using structural equation modelling and clustering (k-means and hierarchical) approaches.

Exploratory Factor Analysis

Exploratory factor analysis is used to identify latent constructs (or 'factors') underlying a set metrics related to schools and to explore the relationships between these metrics and the factors which underpin them.

Confirmatory Factor Analysis

Confirmatory factor analysis is used to explore and validate various models, of increasing levels of complexity, composed of some or all manifest variables and factors. These models are developed using CFA outputs alongside related theory.

Correlation Analysis

Correlation analysis, in the form of correlation matrices, is used to closely examine the associations between the 6 headline SPMs.

Data Used

All data for this study was sourced from the publicly available datasets via the UK Government website "Find and check the performance of schools and colleges in England". This includes data for 3125 mainstream state secondary schools in England for the academic year 2021-2022.

13 variables were selected from the data, including the 6 headline performance measures and 7 metrics of wider school information.

Headline performance measures: Attainment 8; Progress 8; % grade 5+ in English and maths ('Basics'); %EBacc entry; EBacc Average Point Score; Pupil destinations

These are designated as 'headline measures' by the Department for Education to establish their greater importance regarding school accountability and to provide a basis for year-on-year comparison.

Wider school information metrics: Whether the school has a religious character; School gender of entry (mixed or single gender); KS2 Average Point Score; % of pupils with Free School Meals ever in last six years; % English as additional language; % Special educational needs; school admission policy (selective or non-selective).

These are the key items of information about nature of schools and their intake included within the dataset.

Key Findings

The outputs of the EFA suggests that there are 3 latent variables which can be used to predict the 6 SPMs. These latent variables can be conceptualised to be 'Attainment (+ Destinations)' - Attainment 8, Basics, EBacc APS and Destinations, 'Progress' - Progress 8 and 'Curriculum' - % EBacc entry.

'Destinations' has a significantly lower factor loading than the other performance measures, suggesting only a moderate correlation with the underlying factor. While this could itself suggest some degree of breadth in the overall collection of headline performance metrics, it should be noted that Destinations did not load more strongly to another, fourth factor. This could be explained in number of ways. First, the destinations measure is based on the previous academic year's cohort with associated statistical fluctuations. Second, it may suggest the presence of other variables, some of which are beyond the influence of the school (e.g. availability of employment, education and training in the school's locality) not included in the current model. Other sources of data related to these variables could be utilised, allowing them to be incorporated to facilitate more data-informed policy.

Using a CFA/SEM approach, several factors were examined and further developed. These include factors conceptualised as 'Attainment', 'Progress', 'Curriculum', 'Destinations', 'School Character', 'School Context'. These can provide a valuable basis on which to evaluate policy around school accountability, the interpretation of SPMs and the complex meanings and values which they represent.

Correlation analysis revealed that there are extremely strong associations between Attainment 8, Basics and EBacc APS. While this is relatively unsurprising given the common basis of each of these metrics, it is the strength of this correlation which is significant, indicating a substantial narrowing of the concept of 'performance' within the 6 headline SPMs. In turn this may suggest some degree of scope, from the point of view of policy, to replace one or two of these measures with alternative measures, thereby retaining the focus on performance in academic core of subjects, including English and maths, while introducing greater breadth to the concept of performance.

Crucially, more complex models were developed using CFA/SEM which suggest a significant influence on the 6 headline SPMs of factors (e.g. 'school context' and 'school character') which are beyond school influence, bringing into question the use of such metrics alone to judge school performance.

This study forms the basis for further research which will investigate the potential role of digital technologies, including AI, and diverse sources of data to formulate and evaluate more informed intelligence around school performance and context. This can inform advances in education policy related to a less reductionist school accountability system, based on a revised, broader set of principles, e.g. equity and fairness.

Technologies & Analytics

Honest Computing: Achieving demonstrable data lineage and provenance for driving data and process-sensitive policies

Florian **Guitton**[1], Axel **Oehmichen**[1,2,] Étienne **Bossé**[2] and Yike **Guo**[3]

1. *Data Science Institute, Imperial College London, London, United Kingdom*.
2. *Secretarium Ltd, London, United Kingdom*.
3. *Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong*.

Sub.No: DAP-2023-0172 -St2
Full paper is accepted for publication in the Data & Policy journal.

Abstract

Data is the foundation of any scientific, industrial or commercial process. Its journey typically flows from collection to transport, storage, management and processing. While best practices and regulations guide data management and protection, recent events have underscored its vulnerability. Academic research and commercial data handling have been marred by scandals, revealing the brittleness of data management. Data, despite its importance, is susceptible to undue disclosures, leaks, losses, manipulation, or fabrication. These incidents often occur without visibility or accountability, necessitating a systematic structure for safe, honest, and auditable data management.

In this paper, we introduce the concept of Honest Computing as the practice and approach that emphasizes transparency, integrity, and ethical behaviour within the realm of computing and technology. It ensures that computer systems and software operate honestly and reliably without hidden agendas, biases, or unethical practices.
It enables privacy and confidentiality of data and code by design and by default. We also introduce a reference framework to achieve demonstrable data lineage and provenance, contrasting it with Secure Computing, a related but differently orientated form of computing. At its core, Honest Computing leverages Trustless Computing, Confidential Computing, Distributed Computing, Cryptography and AAA security concepts.
Honest Computing opens new ways of creating technology-based processes and workflows which permit the migration of regulatory frameworks for data protection from principle-based approaches to rule-based ones. Addressing use cases in many fields, from AI model protection and ethical layering to digital currency formation for finance and banking, trading, and healthcare, this foundational layer approach can help define new standards for appropriate data custody and processing.

"Smart or not": An Assessment Practice of Customer Service Chatbot from the Chinese Governments Based on Benchmark Testing

Yuting **Huang**[1], Futian **Shao**[2] and Weiyi **Zhang**[3]

1. *School of Government, Peking University*
2. *Laboratory for Government Big Data and Public Policy, Peking University*
3. *Global Development Institute, The University of Manchester*

Sub. No: 904 -St2

Abstract:

Artificial intelligence question-answering has emerged as a crucial application across diverse industries. In the realm of e-government, chatbots function as intelligent customer service agents leveraging natural language processing technology. They autonomously address queries, offering policy consultations and facilitating business transactions. However, the performance of chatbots varies and there is a lack of an objective, replicable and efficient assessment approach in the government service context.

This study presents a performance assessment for customer service chatbots in government, grounded in benchmark testing. The approach evaluates the usability of chatbot responses and assesses their overall performance, presenting a set of comprehensive and quantifiable criteria. This approach serves as a valuable reference for global e-government practices in the domain of AI question-answering.

The assessment framework comprises three primary dimensions: technical capability, knowledge reserve and situational application, with 13 detailed indicators such as understanding ability, associative ability, reasoning ability, etc. A benchmark test set is formulated through an automatic process and manual check, with 1183 questions in total to test 29 provincial-level government customer service chatbots in China. The scoring criteria apply a weighted summation, where the entropy weight method (EWM) is used to get the objective weights of each indicator. Chatbots from the portals of Zhejiang, Heilongjiang and Fujian provinces rank at the top according to final scores. The scores offer a comprehensive reflection of the performance of government customer service.

This study contributes to the field by establishing a robust assessment index and assessment procedure suitable for continuous iterative government customer service chatbots. It also provides a valuable practice in evaluating intelligent government service.

Influence of Covid-19 Pandemic on Population-Level Behavioral Changes: An IoT Based Study in the USA

Jasleen **Kaur**, Arlene **Oetomo**, Vivek **Chauhan** and Plinio **Morita**

1. *University of Waterloo*

Sub. No: 6390 -St2

Abstract

I. INTRODUCTION

The COVID-19 pandemic's restrictions had a significant impact on behavioral markers such as physical, sedentary, and sleep activity. The Behavioral Risk Factor Surveillance System, which is limited by its subjectivity and the quality of the input data, is one of the basic data sources used to assess the current state of the public health surveillance system in the USA. Additionally, there are issues with battery life and data access for data collected by new-generation data sources such as Fitbits. The objective of this research study is to use zero-effort technology (ZET) and IoT-based big data to examine the consequences of the COVID-19 pandemic on population-level behavioral changes in the USA. The objective is to overcome the limitations of the traditional data sources and use the NextGen data sources to examine the consequences of the COVID-19 pandemic on behavioral markers such as physical, sedentary, and sleep activity.

II. METHODS

The study uses the DYD dataset (Donate Your Data) from the Ecobee program, a smart thermostat company, to evaluate the impact of the COVID-19 pandemic on household occupancy patterns and variations in the USA. The size of the USA dataset is approximately 8 TB, over five years (2016–2021) with n = 1,78,706 households captured at five-minute intervals. We implemented the proposed methodology for 470 households in the New Mexico state of the USA. The Microsoft Azure Gen2 data lake is used for the storage of raw data in the form of blobs and the Python notebook in the Azure Databricks is used for data pre-processing, processing, and analysis. The R programming is used for statistical analysis and heatmap visualization. The Gaussian mixture model is used to identify sleep parameters by segmenting the sleep cycle records into different clusters. The quantity of sleep is measured by a motion sensor based on the absence of movement, where an increase in sensor activation indicates a longer duration of household occupancy.

III. RESULTS

The findings show significant changes at the household and population level for the selected behavioral health indicators (sleep time, wake-up time, indoor time, and outdoor time) during the COVID-19 pandemic. People spent more time at home during the COVID-19 pandemic, and their time away from home was significantly reduced. The findings are the: 1) heatmap visualizations at the household level depicting the trend analysis during the Covid-19 pandemic; and 2) statistical analysis to determine test of significance for an average difference in sleep duration, outdoor stay, and indoor stay duration before and during the Covid-19 pandemic.

IV. DISCUSSION & CONCLUSION

Sleep health analysis using IoT data is a novel method of measuring public health indicators objectively using zero-effort technology. These innovative data analytics have the potential to provide real-time insights and alert system activation to monitor, promote, and improve health.

Assessing Human Wellbeing in a Trustworthy AI World: The Complexity of Urban Data

Ayse Giz **Gulnerman**[1] and Florian **Koch**[2]

1. *Land Registry and Cadastre Department, Ankara HBV University, Ankara, Türkiye*
2. *Department of Law and Economics, University of Applied Sciences HTW Berlin, Berlin, Germany*

Sub.No: 2684-St2

Abstract

The urban impact on human well-being is a current and critical issue. Well-being is directly associated with the third goal of the UN Sustainable Development Goals (SDGs), though other SDGs include indicators corresponding to human well-being (Statistics, U. N., 2019). It is important to note that while human well- being in general is associated with several factors such as economic, social, and gender considerations, urban well-being is much more related to urban design, land use planning, and environmental aspects and refers mainly to SDG 11. Nevertheless, these terms are interconnected and mutually influence each other.

Urban well-being is commonly assessed by several indicators, including a green environment, active transport, diversity, density, climate, etc. Since every person experience cities differently, human perception is considered for a granular look at the interaction between the urban environment and individuals. Recent studies indicate a growing use of new forms of urban data, including wearables, smartphones, and urban sensors, in well-being research alongside traditional techniques (Miller et al., 2023; Xu et al., 2023; Reichert et al., 2020). These studies pervasively aim to understand the interaction between urban environments and human mental states.

Although there are several marvelous attempts to retrieve such data and conducting such studies over various urban spaces, most of these attempts remain limited. The most faced problems in these studies are based on technical inadequacy for collecting or accessing spatial data, limited or oriented participation, and privacy issues in collecting and processing such data with AI techniques. These problems limit studies' spatial boundaries, time boundaries and cause spatial, temporal, participation bias.

Moreover, studies carried out solely based on the urban base maps and investigating the correlation between the built environment and mental health issues should be questioned due to ignoring changing environments in human surroundings and their own activity spaces in the urban area but assessing the whole city or neighborhood.

This last concern might be seen as the modifiable areal unit problem (MAUP) (Openshaw, 1981), commonly argued in spatial data science. However, we would like to emphasize that this is more than a 2D or 3D MAUP and indicates multi-dimensional perspectives.

Outcomes of a study has already proved the SDGs indicators on wellbeing do not directly correlate with the subjective well-being of individuals but correlated with "human development index", "economic freedom index", "global competitiveness index" etc. (De Neve & Sachs, 2020). In addition, it is barely said that urban well-being is handled within the indicators lists of the UN SDG 11 or the other SDGs. In this context, we formulate the question of how this complex urban data-requiring problem can be solved and how current data services and technologies support its future. We illustrated the required data framework in Figure 1 and engage current technologies to assist in monitoring human-urban interaction for modeling trustworthy AI systems in the development of high-human-perception-based urban environments.

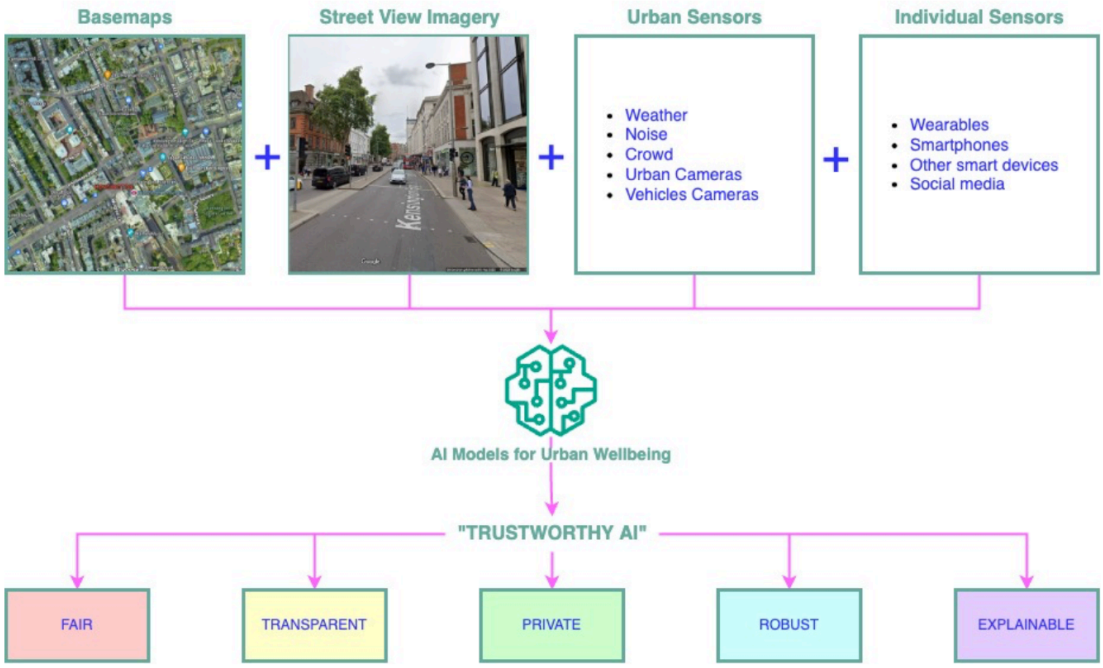

Figure 1. Anatomy of Urban Data on Urban Wellbeing AI Models.

We explore how these systems are allocated within urban spaces and discuss the potential for results to reflect manipulated ideas or lead to the over-monitoring of individuals. Subsequently, we aim to raise the question of how to design such systems to be both useful and high-performing while preserving individual privacy. In conclusion, this discussion highlights the need for a trustworthy AI approach to design urban monitoring systems that balance effectiveness with privacy concerns. Stahl and Leach (2023) mention the importance of ex-ante impact assessment for trustworthy AI to avoid ethical and social

concerns. In this study, we discuss and help to identify the complexity of urban data in the current developments for urban well-being studies. Based on the multi-sourced urban data for the development of AI models for monitoring urban well-being, we discuss if these models have already changed human urban activity spaces. Do urban wellbeing AI models lead to different behavior patterns concerning the use of urban spaces? Is it possible to increase one's individual wellbeing through AI models? If so, should we, or are we going to, base urban policy making on such outcomes?

Furthermore, we expect in the future a growing amount of self-data and contributions from autonomous vehicles, which leads to the creation of different vision-based focal points. This might enhance fair and robust models but also lead to less transparency and an increase in privacy issues.

References

De Neve, J. E., & Sachs, J. D. (2020). Sustainable development and human well-being. World happiness report, 112-127.

Miller, C., Quintana, M., Frei, M., Chua, Y. X., Fu, C., Picchetti, B., ... & Biljecki, F. (2023, November). Introducing the Cool, Quiet City Competition: Predicting Smartwatch-Reported Heat and Noise with Digital Twin Metrics. In Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (pp. 298-299).

Moreno, C., Allam, Z., Chabaud, D., Gall, C., & Pratlong, F. (2021). Introducing the "15-Minute City": Sustainability, resilience and place identity in future post-pandemic cities. Smart Cities, 4(1), 93-111. Openshaw, S. (1981). The modifiable areal unit problem. Quantitative Geography: A British View, 60–69. https://doi.org/info:doi/

Reichert, M., Giurgiu, M., Koch, E. D., Wieland, L. M., Lautenbach, S., Neubauer, A. B., ... & Liao, Y. (2020). Ambulatory assessment for physical activity research: state of the science, best practices and future directions. Psychology of sport and exercise, 50, 101742.

Stahl, B. C., & Leach, T. (2023). Assessing the ethical and social concerns of artificial intelligence in neuroinformatics research: An empirical test of the European Union Assessment List for Trustworthy AI (ALTAI). AI and Ethics, 3(3), 745-767.

Statistics, U. N. (2019). Global indicator framework for the sustainable development goals and targets of the 2030 agenda for sustainable development. Developmental Science and Sustainable Development Goals for Children and Youth, 439.

Xu, J., Liu, N., Polemiti, E., Garcia-Mondragon, L., Tang, J., Liu, X., ... & Schumann, G. (2023). Effects of urban living environments on mental health in adults. Nature Medicine, 1-12.

Predicting the Success of Mobile Money Retail Agents in Ghana: a comparative analysis of well-explored vs less-explored markets using Artificial Intelligence and Machine Learning

Daniel **Osarfo**[1], Peter **Quartey**[1], Agyapomaa **Gyeke-Dako**[2], and Elikplim **Agbloyor K** [2]

1. University of Ghana
2. University of Ghana Business School

Sub. No: 6720-St2

Abstract

The rapid expansion of mobile money services has prompted a heightened interest in understanding the dynamics of mobile money agent operations, particularly in untapped markets or virgin communities. Recent literature emphasizes the role of mobile money agents as pivotal actors in extending financial services to underserved populations (Demirgüç-Kunt et al., 2018; GSMA, 2021) and in sustaining the development of financial systems in developing countries (Donovan, 2012; Senyo, Karanasios, Gozman & Baba, 2022). The existing literature also presents a complex and often inconclusive picture, with contradictory findings and limited predictive power, relying on small scale studies, static models and qualitative approaches.

This underscores a need for further inquiry into decision-making processes of mobile money agents, especially in choosing to enter, succeed, or exit a market, be it an explored or virgin market. We will employ artificial intelligence and machine learning, due to their ability to handle complex and non-linear relationships within datasets when risks need to be modeled in the analysis (Jack & Suri, 2014; Amini, et al, 2021). Data will be sought from the leading mobile network operator MTN Ghana. The data will include information on agent operation since 2016, the year after Ghana opted for an MNO-led mobile money regime. Such an expansive data will enable us account for the various policy changes that have occurred in the industry.

Policy & Literacy for Data

Identifying stakeholder motivations in normative AI governance: a systematic literature review for research guidance

Frederic **Heymans**[1] and Rob **Heyman**[1]

1. *imec-SMIT, VUB, Brussels, Belgium*

Sub. No: DAP-2023-0163 -St3
Full paper is accepted for publication in the Data & Policy journal.

Abstract
Ethical guidelines and policy documents destined to guide AI innovations have been heralded as the solution to guard us against harmful effects or to increase public value. However, these guidelines and policy documents face persistent challenges. While these documents are often criticized for their abstraction and disconnection from real-world contexts, it also occurs that stakeholders may influence them for political or strategic reasons. While this last issue is frequently acknowledged, there is seldom a means or a method provided to explore it. To address this gap, the paper employs a combination of social constructivism and science & technology studies perspectives, along with desk research, to investigate whether prior research has examined the influence of stakeholder interests, strategies, or agendas on guidelines and policy documents. The study contributes to the discourse on AI governance by proposing a theoretical framework and methodologies to better analyse this underexplored area, aiming to enhance comprehension of the policymaking process within the rapidly evolving AI landscape. The findings underscore the need for a critical evaluation of the methodologies found and a further exploration of their utility. Additionally, the results aim to stimulate ongoing critical debates on this subject.

Commons for the Commons: Climate Action in the Amazon Rainforest through AI and Data

Carolina **Banda**[1] and Germán **Johannsen** [1]

1. *Max Planck Institute for Innovation and Competition and University of Munich*

Sub. No: 1340-St3

Abstract

Humanity's most pressing crisis, climate change, calls for a paradigm shift from the old market-state dichotomy towards a mission-oriented action, including all relevant actors. By adopting a Law, Tech, & Society approach, we propose a legal theory on the governance of the data commons, aligned with UN Sustainable Development Goal N°13 on climate action. This is what we call the **Commons for the Commons**. This theoretical framework is tested against four selected use cases which address the problem of deforestation in the Brazilian Amazon through the use of AI and data technologies. These use cases are 1) the prosecution of illegal deforestation, 2) product traceability, 3) the transition to a bioeconomy, and 4) big-tech data commons initiatives. We believe that clear data governance principles are needed to unlock the potential of these initiatives. The ultimate goal of our research is to make more environmental data available to aid policy decision-making, improve transparency in the data value chain, and increase accountability for the impact of deforestation in the Amazon rainforest.

- **Background, hypothesis and methodology**

Empirical evidence shows that AI and data are being used in business and non-profit initiatives to combat deforestation in the Brazilian Amazon. However, these initiatives face challenges such as scalability, lack of stakeholder coordination, and unclear data governance.

A data commons governance can leverage AI and data potential to fight deforestation in the Amazon, in turn helping to achieve the UN Sustainable Development Goal on Climate Action (SDG 13).

Our methodology includes developing a legal theory, identifying use cases through desk research, gathering insights from semi-structured stakeholder interviews, analysing this information within the Brazilian data legal and public policy framework, and recommending governance principles and data institutions, such as data trusts or collaboratives, tailored to each use case.

- **Developing a Legal Theory**

We contribute to creating a legal theory on the governance of data commons, particularly for fighting climate change. We adopt a Law, Tech & Society approach aligned with SDG N°13 and its interplay with other SDGs. We support our theory on Ostrom's work, which has provided a valuable framework in relation to community participation in economic

development. Additionally, Purtova et al. discuss a data governance framework that highlights the relevance of sustainable governance of data commons itself, and a political-moral dimension for data governance. In the context of climate change and the Amazon, we identify that this dimension for us are the SDGs. Hence, our research aims to explore two intertwined dimensions of sustainability: an internal one, focused on the sustainability of the data commons, and an external one, focused on the social purpose given to data governance, i.e., climate action.

A main criticism of Commons is its lack of scalability. As Mazzucato points out, this is because the common perspective is seen as a counterpoint to a weak or captured state. In contrast, Mazzucato suggests that the public sector should play a more participatory role, setting ambitious goals to address the grand challenges of our time, such as climate change, by promoting collective action among multiple stakeholders. This paradigm shift promotes an economy that generates value from the activities of civil society, industrial policy and markets.

Our research draws on this new vision of the state's role and places it in the context of climate action and the potential of data commons.

- **Empirical analysis**

The authors of this paper are part of a bigger project called 'Data Governance in Emerging Economies to Achieve SDGs'. As part of this project, we organised a workshop with different stakeholders in Sao Paulo in December 2022. Here, we gained insights from businesses, academia, and civil society. Before and after the workshop, we conducted semi-structured interviews and desk research, identifying four use cases.

- **Legal analysis**

In Brazil, a Data Protection Law (LGPDP) regulates personal data. However, environmental data such as deforestation level falls outside this framework. Data sharing is happening despite the absence of a specific law for non-personal data, such as in the case of the environment. The absence of a clear data law framework represents specific legal challenges for each use case:

1. **Data+AI to support the prosecution of illegal deforestation:** There are different initiatives by NGOs and the government, but there is insufficient coordination to create an efficient system to enforce sound environmental laws.

2. **Data+AI to improve the traceability of product origin along the value chain:** While a new EU Regulation seeks to avoid the placement of certain commodities contributing to deforestation, there is no such law in Brazil. Yet, data and AI platforms exist that map the origin of products and increase transparency along the value chain.

3. **Data+AI to transition from a monoculture to a bioeconomy:** some private and NGOs initiatives are promoting bioeconomy innovation, but there is still a 'first mover disadvantage' related to this transition, thus maintaining the industrial status quo.

4. **Big-tech & data commons:** Google's recent "philanthropic" initiative on data commons might raise concerns about its profit-oriented agenda regarding data use, competition, and innovation risks due to the control of data and AI resources.

- **Proposal of legal and policy principles**

The coordination problem could be a result of the absence of an adequate legal framework that promotes collective initiatives for climate action (e.g., the definition of standards to lower transaction costs). There is also legal uncertainty regarding the legality of data-sharing agreements (e.g., competition law on horizontal agreements between competitors). Against this backdrop, we propose a data commons governance with three lines of action: (i) designing legal tools based on data governance principles to foster data commons for climate action. One example of these principles is 'use-case by design'; (ii) developing guidance about the legal interpretation of AI and data sharing agreements in the context of data collaboratives; and finally (iii) for escalating data collaboratives, the state shall have a more active role and be the orchestrator and facilitator of data governance frameworks towards the common good.

Leveraging data ecosystems to address climate challenges: an urban perspective

Natalia **Oprea**[1] and Charlotte **van Ooijen**[2]

1. *The Lisbon Council, SDA Bocconi School of Management*
2. *CvanO - Digital Government Research and Advice*

Sub. No: 8347-St3

Abstract

The impact of climate change is most strongly felt at the local level. As cities are confronted with increased air pollution, floodings and heat islands, local governments are looking for innovative ways to address such challenges. Data and emerging technologies like artificial intelligence may be leveraged to better understand climate challenges and devise policy solutions to tackle them. Nevertheless, the transformative potential of data remains largely unutilized, with both public and private actors striving to realise the value of data (Fussell 2023).

To unlock the potential of data-driven innovation, the combination of various data sources and the collaboration of different organisations are needed (Ryazanova et al. 2016). This is especially true for local governments, which notoriously deal with limited data and technological resources than their national counterparts. The distributed nature of actors and their sometimes-conflicting aims when collecting and sharing data, however, pose challenges by themselves. Data ecosystems have been suggested as fruitful environments for the collaboration between autonomous actors to explore data (Oliveira et al. 2019).

This paper contributes to the research field of data ecosystems (Liva et al. 2023), analysing the conditions for realising an urban data space to address climate challenges. Data spaces, launched by the European Strategy for Data (EC, 2020), can be considered a manifestation of the idea of data ecosystems, with their true innovation lying in an emphasis on participants rather than technological tools. Using a sociotechnical system perspective, the emergence of urban green data spaces in four middle-sized European cities of Ferrara, Graz, Leuven and Zaragoza is studied. A qualitative methodology approach was adopted to collect and analyse data coming from semi-structured interviews and document research. The analysis explores the components and the combined effect of policies, stakeholder needs, coordination mechanisms and technology on the creation of data spaces.

The results describe the difficulty to envision an urban data space with both public authorities, private actors and citizens as contributors to the data value chain. Local authorities mostly focus their efforts on enhancing coordination across municipal departments and integrating available data into their information systems. Thus,

unsurprisingly, interviewees strongly emphasised technical needs to address limitations in interoperability and standardisation, and the diffusion of open-source solutions. Climate policy itself adds another layer of complexity as efforts in this area are not always approached holistically, with initiatives spanning different strategies or municipal offices and departments. At the same time, interviewees demonstrated an awareness that tackling climate issues cannot be a one-sided enterprise and that multi-stakeholder engagement is crucial. In fact, an overarching demand of city administration respondents was to better coordinate citizen involvement and secure the participation of other stakeholders, such as academic groups, with active roles in data exploration and visualisation.

The concept and realisation of a data space to address climate challenges is a fairly uncharted area for local authorities. A first step in building an urban green deal data space, according to the interviewees, is to generate awareness for all relevant stakeholders about the importance of data and data sharing. In that same direction, stakeholders' interaction and collaboration should be facilitated by harmonised methods of data collection, use and management. Finally, joint efforts should be climate-challenge driven, meaning that individual stakeholders'' objectives should be aligned with climate targets driving the creation of the data space.

References:

European Commission (2020) A European Strategy to Data. Brussels: European Commission.

Fussell, C. (2023, February 8). Why we struggle to realise the value of data. https://doi.org/10.31235/osf.io/u8zcx

Liva G, Micheli M, Schade S, Kotsev A, Gori M and Codagnone C (2023). City data ecosystems between theory and practice: A qualitative exploratory study in seven European cities. Data & Policy, 5: e17.

Oliveira M, Barros Lima G and Farias Lóscio B (2019) Investigations into data ecosystems: A systematic mapping study. Knowledge and Information System 61, 589–630.

Ryazanova, O., P´etercs´ak, R., Heaphy, L., Connolly, N., & Donnellan, B. (2016). Perception of value in public-private ecosystems: Transforming the Dublin docklands through smart technologies (pp. 1–10).

Datathon on Gender and Racial Inequalities in Public Service: an innovative data literacy experience from Brazil.

Carolina **Coppetti**
ENAP (École nationale d'administration publique)

Sub. No: 7042 -St3

Abstract

Introduction

This article aims to present an innovative experience on data literacy developed by the National School of Public Administration from Brazil during the year of 2023. A series of Datathon events was idealized to not only shed light on existing gender and racial inequalities in public service but also to show the importance of data literacy and diversity awareness in policymaking.

Methodology

In total, 80 diverse participants from all over the country engaged in a hands-on in-place experience, working with secured datasets on Brazilian federal civil servants information, each time during a whole week. There were two editions of this program so far, one focused on gender and another on racial inequalities, even though the participants were encouraged to complexify their analysis through an intersectional approach. With guided workshops and specialized mentorship on gender studies, racial studies, design thinking and data, teams were encouraged to employ statistical analysis and data visualization techniques to propose innovative solutions to inequalities in public service.

Results

The outcomes of these two Datathon experiences included, first of all, a gain in the ability of participants to use data as a relevant tool of insights. Participants were motivated by understanding how to solve the complex challenges faced by most of them, underrepresented groups within the public service sector, specially on assuming leadership positions and feeling recognized as valuable civil servants. During each week, it became meaningful for them to explore data analysis and visualization as new tools. Through collaborative efforts, all the 16 teams produced actionable recommendations aimed at addressing systemic issues and fostering inclusivity in public service.

Conclusion

The Brazilian public service experience demonstrated the capacity of Datathon events to act as a catalyst for bridging the gap between data literacy and social justice. By developing a meaningful data and diversity literacy ambiance, we empowered individuals to become advocates for equitable practices within public service. Several participants continued their trajectory on data education. The ripple effect of this initiative extends beyond the event itself, contributing to a more data-based and diversity-considering approach to policymaking and organizational practices.

An Analysis of the Lifecycle of Generative Artificial Intelligence in Industrial Settings: Implications for Governing Risks and Responsibilities among Stakeholders.

Hillary **Giam**[1] and Masaru **Yarime**[1]

1. *The Hong Kong University of Science and Technology*

Sub. No: 9833 -St3

Abstract

The reported transformative potential of integrating Generative Artificial Intelligence (GenAI) across various sectoral domains has galvanized the collective imagination and anxiety across industry and policymakers alike (Blackman, 2023). The estimated business value and productivity gains of GenAI have encouraged its integration into different sectoral use cases (McKinsey & Company, 2023). However, the deeper embeddedness of GenAI in commercial settings also amplifies pre-existing concerns wherein unpredictable and untraceable outcomes compound a lack of interpretability, transparency, and accountability on the content produced (Abusitta et al., 2019; Wang et al., 2023). This becomes exponentially damaging when generative technologies are deployed to inform socio-economic decisions in financial services or healthcare diagnosis (Harrer, 2023).

From the policymaker's perspective, GenAI's rapid integration and growing embeddedness in the commercial sphere hold policy implications across multiple domains. The developing regulatory landscape, coupled with fuzzy boundaries on the legal processing of data and copyright, results in an open question of what guardrails are truly needed for responsible GenAI.

The research question aims to examine the lifecycle of the design, development, and deployment of GenAI in commercial settings. Using a case study approach and semi-structured interviews, our research clarifies the interactions between actors and the transfer of sources of data throughout the GenAI lifecycle in four sectors. For effective model governance, a holistic examination of the GenAI lifestyle remains pertinent to identify the sources of risk and responsibility within and across industries. Our findings will inform the institutional design of model governance practices as policymakers strive toward responsible GenAI.

The paper anchors on risk and regulation literature to systematically identify, evaluate, and control for sources of risk in the commercial design, development, and deployment of GenAI technologies (Lodge & Wegrich, 2012; Steimers & Schneider, 2022). Risk identification and evaluation delineates the extent of damage and probability of occurrences of potential hazards. For GenAI in commercial settings, the interaction between the source of data used, sector-specific characteristics, the type of foundational model used, i.e., open-sourced versus pre-trained models, the developers of the foundational models versus sector-specific models, and the communication between developers and end-users, alters the risk scorecard for each industry. The interplay between these variables remains increasingly essential to understanding the evolving nature of the sources of risk throughout the GenAI lifecycle.

Risk control measures focus on identifying and implementing governance options to manage the sources of risk. It dictates an allocation or sharing of responsibility to address the identified risks, either through collaboration or competition among the actors. Broadly, three governance approaches are used as risk control measures for the responsible design, development, and deployment of AI technology (Abbott & Snidal, 2000; Cath, 2018). Principle-based governance leverages soft laws and guidelines as risk control measures. Rule-based governance includes regulatory standards or legislations to ensure compliance, accountability, and transparency from regulated entities to manage the identified risks. Adaptive regulation offers responsive approaches such as self-regulation, co-regulation, or performance-based mechanisms to ensure regulated entities address the sources of risk.

Once risk control measures have been implemented, residual risks are re-evaluated with a risk-benefit analysis to determine the acceptability of the overall residual risk. Evaluating the effectiveness of risk control measures informs policymakers on the optimal approach towards designing model governance best practices. Despite the limited specificity of AI governance for generative models, the insights provide a baseline to establish risk control measures for GenAI in commercial settings (Ayling & Chapman, 2022).

This paper conducts case studies to unpack the commercial application of GenAI in the media and entertainment, manufacturing, healthcare, and financial services sectors. The selected sectors capture a comprehensive spectrum of GenAI's primary six modalities of use in audio, code, data analysis, image, text, and video (Kanbach et al., 2023; Seawright & Gerring, 2008). Semi-structured interviews with industry experts and key stakeholders are conducted to validate, inform, and uncover how each sector designs, develops and deploys GenAI models. The interviews provide insight into the prevailing challenges and future direction of GenAI's sectoral integration.

From the qualitative findings, a thematic analysis synthesizes commonalities and differences between each sector's development and deployment of GenAI. The comparison is useful for policymakers to obtain a more nuanced understanding of sector-specific versus overlapping concerns and risks associated with the commercial application of GenAI. It also shapes how policymakers tailor their policy toolkits to achieve model governance in managing the risk of generative models within the sector and across the board.

From the case studies and interviews, the paper aims to illustrate how GenAI is designed, developed, and deployed in commercial settings within the four sectors. In identifying the interactions between actors and the transfer of sources of data, the findings seek to provide policymakers with an understanding of where and how potential risks emerge, are transferred, and amplified throughout the GenAI lifecycle. It formatively establishes a structure of responsibility and accountability for each actor in each stage of the GenAI lifecycle. From the thematic analysis, the paper expects to draw lines of convergence and dichotomy between how each sector approaches GenAI integration. It supports policymakers in developing the model governance landscape, determining whether a universal or industry-tailored approach is more appropriate for GenAI guidance and regulation.

Generative models have existed since the 1950s, with the development landscape undergoing numerous summers and winters, representing fluctuating interest and funding within the field (Goodfellow et al., 2020). Despite its re-emerging hype, the GenAI regulatory landscape remains nascent as policymakers and the industry seek to harness its transformative potential while managing the accompanying risks. Discussions have centred around 'who' should be involved in risk identification, evaluation, and control process, and 'what' aspects within the GenAI lifecycle necessitate regulatory oversight. The paper aims to clarify how GenAI is developed and deployed in commercial settings to provide a foundation to answer these core questions of responsibility and accountability. Then, we can explore building consensus between policymakers and the industry on GenAI model governance.

References

Abbott, K. W., & Snidal, D. (2000). Hard and Soft Law in International Governance. International Organization, 54(3), 421–456. https://doi.org/10.1162/002081800551280

Abusitta, A., Aïmeur, E., & Wahab, O. A. (2019). Generative Adversarial Networks for Mitigating Biases in Machine Learning Systems (arXiv:1905.09972). arXiv. https://doi.org/10.48550/arXiv.1905.09972

Ayling, J., & Chapman, A. (2022). Putting AI ethics to work: Are the tools fit for purpose? AI and Ethics, 2(3), 405–429. https://doi.org/10.1007/s43681-021-00084-x

Blackman, R. (2023, August 14). Generative AI-nxiety. Harvard Business Review. https://hbr.org/2023/08/generative-ai-nxiety

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences, 376(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139–144. https://doi.org/10.1145/3422622

Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. eBioMedicine, 90. https://doi.org/10.1016/j.ebiom.2023.104512

Kanbach, D. K., Heiduk, L., Blueher, G., Schreiter, M., & Lahmann, A. (2023). The GenAI is out of the bottle: Generative artificial intelligence from a business model innovation perspective. Review of Managerial Science. https://doi.org/10.1007/s11846-023-00696-z4

Lodge, M., & Wegrich, K. (2012). Managing Regulation: Regulatory Analysis, Politics and Policy. Bloomsbury Publishing Plc. http://ebookcentral.proquest.com/lib/hkust-ebooks/detail.action?docID=4763590

McKinsey & Company. (2023). Generative AI Global Automation. http://ceros.mckinsey.com/generative-ai-global-automation

Seawright, J., & Gerring, J. (2008). Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. Political Research Quarterly, 61(2), 294–308.

Steimers, A., & Schneider, M. (2022). Sources of Risk of AI Systems. International Journal of Environmental Research and Public Health, 19(6), 3641. https://doi.org/10.3390/ijerph19063641

Wang, S., Zhao, C., Huang, L., Li, Y., & Li, R. (2023). Current status, application, and challenges of the interpretability of generative adversarial network models. Computational Intelligence, 39(2), 283–314. https://doi.org/10.1111/coin.12564

Drafting an 'AI Policy' for organizational use: Development Gateway's experience

Beverley **Hatcher-Mbu**[1], Tom **Orrell**[1] and Jacob **Leiken** [2]

1. Development Gateway: An IREX Venture
2. NYU Law School

Sub. No: 4022 -St3

Abstract

Policy Question

With the rise of generative AI, organizations of all sizes aim to take advantage of the AI revolution but do not know how to do it safely. With concerns around intellectual property, privacy, and confidentiality, potential users of generative AI services often cannot make informed, ethical decisions about which services are both fit for purpose and safe to use. This is exacerbated for small NGOs, without the resources to carefully audit, or the bargaining power to contract evenly with AI service providers. Yet many have questions about whether they can use new services to summarize documents, generate code, create images, and more.

So, we asked ourselves: how can we write and follow a forward-looking internal AI policy, given the considerable information and power asymmetry present in this space?

Methodology

First, we needed to understand how people were using or wanted to use AI tools across our teams. These teams included project managers and software developers with a broad range of job responsibilities. We surveyed staff across 6 countries and generated a working list of the tools they flagged. Our software team researched the tools and their user agreements to help understand the risks and opportunities of our current usage.

This research informed a draft assessment rubric, which could be applied to any new service as needed. We developed a cross-team working group to revise the rubric. The group met over several weeks to develop and refine an approach grounded in the existing use cases we identified from across our teams, and prepare for new use cases to come. Once the working group was satisfied with the draft, we opened the document to broader feedback across the organization.

We sought feedback from all levels of seniority, from entry-level staff to directors. This was essential because we found, anecdotally, that entry-level staff were using particular tools more heavily, were more invested in automating some of their tasks, and were looking more actively for emerging AI tools that could fit these needs.

Key Findings

The final version of our "AI Internal Policy" is based on a two-pronged system: both the tool and use case must be approved. Both lists are provided in a policy guidance document with rationale and additional guidelines.

Most notably, the guidance includes a checklist for the analysis of new tools. The checklist directs staff to think through the possible ramifications of their use, as well as analyze the tool's terms and conditions. Even with review of use and T&Cs, the safety of the tool is not always clear cut. Team members are encouraged to share new use cases with the cross team working group to support the risk assessment review. If the tool does not raise any red flags based on the checklist, the team member can begin using the tool, notifying the working group of the new use case.

Most name-brand AI tools have been approved; however, tools are rejected if their data sharing policies are too lenient. One tool was rejected for having terms and conditions which do not include a data protection clause. Many use cases that are approved come with restrictions. For example, software developers can only input code into AI tools if the code is from an open-source repository. This forecloses the possibility of proprietary code leakage. Meeting transcriptions can only be generated by AI with verbal consent from all parties, to preserve client trust and honor personal privacy. Some use cases are banned entirely: no AI-produced images can be shared externally, to minimize legal uncertainty around copyright.

Reflecting the difficulty presented by information asymmetry, the working group discussed the following aspects at length to balance the need for practicality (and thoroughness):

- Reviewing the terms and conditions of potential tools could involve extensive, in-depth analysis by dedicated experts. They would evaluate a tool based on its storage and deletion practices, and whether user input data is used to train future iterations of AI models. To address this, our checklist adopts a simplified approach to provide clear, easy-to-follow guidance, to ensure that team members can reliably comply without becoming overwhelmed.
- Many reports identify bias in widely used AI models. We do not want to feed into bias or incorporate bias into our work. Yet, this is an extremely difficult problem to address, with whole companies dedicated to measuring and eliminating bias in AI models. Our team has only limited time and expertise to identify serious concerns. Our guidance notes this issue and requires instances of bias to be reported to the AI working group. As we are often unable to affect these models meaningfully, the most effective way for us to address bias in AI as practitioners is to minimize and supervise its use and ban rogue tools as they are identified.
- We do not want staff to use an AI tool to perform most of their job tasks. Especially in its current state, the output of an AI model must be reviewed in detail before it is used in a professional context (for example to avoid the impact of algorithmic hallucination and misinformation). Rather than limiting an individual's usage, the policy focuses on building healthy habits around intentional usage, thereby avoiding the pitfall of trying to "police" team behavior across multiple countries and legal regimes.

Our presentation will discuss the decision-making process behind each provision and those which were considered and not included, in order to support similarly situated organizations and individuals in their efforts to use AI tools thoughtfully and responsibly.

Ethics, Equity & Trustworthiness

AI-assisted pre-screening of biomedical research proposals: ethical considerations and the pilot case of "la Caixa" Foundation

Carla Carbonell **Cortés**[1], César **Parra-Rojas**[2], Albert Pérez-**Lozano**[3], Francesca **Arcara**[2], Sarasuadi Vargas-**Sánchez**[2], Raquel **Fernández-Montenegro**[3], David **Casado-Marín**[1], Bernardo **Rondelli**[2] and Ignasi **López-Verdeguer**[1]

1. *Area of Partnerships with Research and Health Institutions, "la Caixa" Foundation, Barcelona, Spain*
2. *SIRIS Lab, Research Division of SIRIS Academic, Barcelona, Spain*
3. *Analytics & Artificial Intelligence, IThinkUPC S.L.U., Barcelona, Spain*

Sub. No: DAP-2023-0159 -St4
Full paper is accepted for publication in the Data & Policy journal.

Abstract

The "la Caixa" Foundation has been experimenting with AI-assisted decision-making geared towards alleviating the administrative burden associated with the evaluation pipeline of its flagship funding programme, piloting an algorithm to detect immature project proposals before they reach the peer-review stage, and suggest their removal from the selection process to a human overseer. In this paper, we explore existing uses of AI by publishers and research funding organisations to automate their selection pipelines, in addition to analysing the conditions under which the focal case corresponds to a responsible use of AI and the extent to which these conditions are met by the current implementation, highlighting challenges and areas of improvement.

A Feminist Framework for Urban AI Governance: Addressing Challenges for Public-Private Partnerships

Laine **McCrory** [1,2]

1. *Department of Communication and Culture, Toronto Metropolitan University, Toronto, Canada*
2. *Department of Communication and Media Studies, York University, Toronto, Canada*

Sub. No: DAP-2023-0175 -St4
Full paper is accepted for publication in the Data & Policy journal.

Abstract

This analysis provides a critical account of AI governance in the modern "smart city" through a feminist lens. Evaluating the case of Sidewalk Labs' Quayside Project - a smart city development that was to be implemented in Toronto, Canada - it is argued that public-private partnerships can create harmful impacts when corporate actors seek to establish new 'rules of the game' regarding data regulation. While the Quayside project was eventually abandoned in 2020, it demonstrates key observations for the state of urban algorithmic governance both within Canada and internationally. Articulating the need for a revitalized and participatory smart city governance program prioritzes meaningful engagement in the forms of transparency and accountability measures. Taking a feminist lens, it argues for a two-pronged approach to : integrating collective engagement from the outset in the design process, and ensuring the civilian data protection through a robust, rights-based privacy regulation strategy. Engaging with feminist theories of intersectionality in relation to technology and data collection, this framework articulates the need to understand the broader histories of social marginalisation when implementing governance strategies regarding artificial intelligence in cities.

Towards Fairer AI: A Visual Synthesis of Bias Mitigation Tools and Training Frameworks

Alenka **Guček**[1], Tanja Zdolšek **Draksler**[1], Matej **Kovacic**[1], Andreas **Karabetian**[2], Konstantinos **Mavrogiorgos**[2] and George **Manias**[2]

1. *Institut Jožef Stefan, Ljubljana, Slovenia*
2. *University of Piraeus, Pireas, Attica Greece*

Sub. No: DAP-2023-0161 -St4

Abstract

In the ever-evolving landscape of artificial intelligence (AI), mitigating bias is a critical imperative for ensuring equitable and ethical decision-making. This paper introduces a comprehensive Bias Detector Toolkit that serves as a visual catalogue synthesizing various bias mitigation tools for bias detection in AI. The toolkit is designed to empower general public, developers, researchers, and practitioners with a nuanced understanding of the diverse range of tools available for detecting and mitigating biases in AI systems. The toolkit is unveiled through scrolytelling, a narrative method intertwining scrolling and storytelling. Using vivid visual metaphors, we navigate the intricacies of bias, emphasizing the need for robust mitigation strategies. Providing real life examples, all relevant stakeholders can get a better grip of importance for bias mitigation. Structured as a dynamic visual synthesis, the catalogue offers a comprehensive overview of bias mitigation tools, categorized by functionalities and applications. Serving as a visual catalogue, it facilitates exploration, comparison, and informed tool selection, fostering a more effective approach to bias mitigation. In the final segment, we shift focus to the development of a learning framework that aims to empower AI practitioners with the knowledge and skills necessary for implementing bias mitigation strategies. By combining theoretical insights with hands-on practical exercises, the learning framework addresses the educational gap in AI bias mitigation, fostering a community of practitioners. This holistic approach—from scrolytelling to bias detection tools and learning frameworks—forms a unified strategy towards advancing the field of fair and democratic AI.

The Dark Side of Large Language Models: Legal and Ethical Challenges from Stochastic Parrots and Hallucination.

Zihao **Li**
University of Glasgow & Stanford University

Sub. No: 1471 -St4

Abstract

With the launch of ChatGPT, Large Language Models (LLMs) have been revolutionizing various aspects of society, rapidly altering the way we think, create and live. For instance, the GPT integration in Bing has altered our approach to online searching. While nascent LLMs have many advantages, new legal and ethical risks are also emerging, particularly exemplified by the phenomena of "stochastic parrots" and "hallucination." These terms describe the tendency of LLMs to generate unverified information and to replicate patterns from training data without true understanding the context and content. Hallucination in LLMs can lead to the dissemination of confidently presented but factually incorrect information, posing significant risks in contexts where authenticity is critical, such as healthcare and legal advice. Similarly, stochastic parrots, by merely echoing training data, can perpetuate biases and stereotypes, potentially reinforcing harmful social prejudices and misleading decision-making processes. Additionally, subtle inaccuracies, oversimplifications, or biased responses passed off as truth in a confident tone pose a substantial risk in research, science communication, and education, as they can mislead both experts and non-experts, undermining the integrity of scientific inquiry and knowledge dissemination.

The European Union (EU), as a frontrunner in AI regulation, has focused on the regulation of AI models. However, these risks posed by LLMs are likely to be underestimated by the emerging EU regulatory paradigm. The efficacy of existing EU regulatory framework, such as the proposed AI Act (AIA), Digital Service Act (DSA), and Digital Markets Act (DMA), is questionable to adequately address the intricate risks of hallucination and stochastic parrots associated with general-purpose LLMs. This article delves into these challenges, assessing their implications within the legal and ethical domains and scrutinizing the adequacy of the EU's regulatory framework in addressing these risks posed by general-purpose LLMs.

Research questions:

Based on above context, this article addresses three principal research questions: (i) What are the distinct legal and ethical risks posed by the phenomenon of "stochastic parrots" and "hallucination" in LLMs? (ii) Is the current EU AI regulatory framework adequate in addressing such risks posed by general-purpose LLMs? (iii) What advancements are required in the EU AI regulatory paradigm to effectively mitigate the risks associated with stochastic parrots and hallucination?

Method:
This research employs an interdisciplinary approach, incorporating both technical and legal doctrinal analysis. The technical analysis involves a systematic examination of the operational mechanics and algorithmic architecture of LLMs, focusing on phenomena such as "stochastic parrots" and "hallucination." This is complemented by a legal doctrinal method, which entails a detailed exploration of existing legal standards, regulations, case law, academic literature, and expert opinions. This method involves an in-depth examination and interpretation of legislations and policy, aiming to establish a normative framework for understanding and addressing the challenges posed by LLMs. This interdisciplinary approach provides a solid foundation for the article's regulatory recommendations, designed to inform data scientists, AI developers, and policymakers.
Key findings:
The key findings and contributions of this article are multi-faceted. It first taxonomize the phenomenon of hallucination and stochastic parrots in LLMs and explores how it impacts the authenticity, reliability and trustworthiness of AI-generated information. This research also differentiates the legal implications of stochastic parrots and hallucination, demonstrating their effect on the credibility and trustworthiness of information, especially in sensitive sectors like healthcare and law. However, merely improving the accuracy of the models through new data and algorithms is insufficient, because the more accurate the model is, the more users will rely on it, and thus be tempted not to verify the answers, leading to greater risk when stochastic parrots and hallucinations appear. This situation, where an increase in accuracy leads to higher reliance and potential risks, can be described as the 'accuracy paradox'.
This article secondly identifies deficiencies in the current EU regulatory framework, including the AI Act, DSA, and DMA. These regulations, while groundbreaking, are found lacking in effectively addressing the unique challenges presented by LLMs, primarily due to their limited scope in categorizing and regulating these models. It points out that the existing regulatory approach, exemplified by DSA and DMA, is a consequence of the current platform-as-a-service (PaaS) business model. However, once the business model shifts to AI model-as-a-service (MaaS), this regulatory framework is likely to become nugatory, as the platform does not fully control the processing logic and output of the algorithmic model. Regarding the AI Act, although the recent debate concentrates on the obligations of generative and general-purpose AI, output moderation of LLMs is largely underdiscussed. Only the transparency obligation is insufficient to tackle the accuracy paradox issue.

Therefore, based on above discussion, this research proposes an innovative solution combining content and output moderation across the LLM deployment value chain, with a focus on enhancing data quality. It aims to tackle the root causes of unreliable and untrustworthy AI outputs and inherent biases in LLMs. It suggests that for sensitive areas, LLMs should be designed to guide users towards authoritative sources, underscoring the importance of user verification of AI-generated content. Moreover, this article advocates a paradigm shift in the EU AI regulatory framework, transitioning from a risk-based approach to trustworthiness-based approach. It implies that AI regulation should move from focusing solely on transparency to ensuring the reliability, explainability, and traceability of both AI training data and AI-generated information, enabling the EU to leverage the benefits of LLMs while mitigating potential risks.

Responsible AI Mechanisms in Public Sector Organizations: A Realist Synthesis Review.

Ana **Gagua**[1], Haiko van der **Voort**[1], Nihit **Goyal**[1] and Alexander **Verbraeck**[1]

1. *TU Delft, Faculty of Technology, Policy and Management*

Sub. No. 8637 -St4

Abstract

This study explores the mechanisms public sector organizations use to govern AI responsibly, examining how these mechanisms vary across different contexts. Using a realist synthesis approach, the research aims to identify practical, context-specific mechanisms and related outcomes to bridge the gap between ethical AI principles and real-world implementation in public sector organizations.

1- Introduction

Responsible AI and related principles such as transparency, accountability, fairness and explainability has gained widespread attention in recent years (Dignum, 2019; Zhu, 2019). RAI is considered a crucial step in addressing the criticisms that AI technologies are often biased, opaque, and unfair.

This focus becomes even more critical in government organizations due to their responsibility to uphold the highest public value standards and their high influence on citizens. The government sector worldwide increasingly adopts AI technologies to enhance government effectiveness and efficiency (Bertot et al., 2016; Zuiderwijk, Chen, & Salem, 2023), which is accompanied by the challenge of integrating AI in a manner that meets societal values and policy standards.

However, despite creation of over 70 documents outlining ethical principles or frameworks by different stakeholders from various sectors (Floridi & Cowls, 2019; Jobin et al., 2019), application of these principles in real-life settings remains ambiguous. There is a significant gap between what ethical frameworks aim to achieve and what is practically implemented (Mittelstadt, 2019; Morley et al., 2020; Hagendorff, 2020). Specifically, mechanisms leading to RAI in different public sector contexts remain unclear. We address this gap by posing the question:

- which mechanisms might lead to responsible AI in the public sector organizations?
- In what organization-related contexts have these mechanisms been researched in?
- In what academic context (e.g., research disciplines, geographic regions) have these mechanisms been studied?

1.1 Research Methodology

Using a realist synthesis approach, we aim to provide a comprehensive and empirically grounded understanding of effective AI governance mechanisms. This approach is particularly suited for understanding complex interventions and their contexts. It focuses on identifying what works, for whom, in what circumstances, and why (Pawson, 2005). This is essential for AI governance and ethics, where context-specific factors play a significant role (Wong, 2013). While systematic reviews are excellent for aggregating evidence and assessing effectiveness, they often lack the depth needed to explore the contextual nuances and mechanisms underlying AI governance (Pawson, 2005). For policy-related areas, there is increasing interest in realist synthesis as an alternative method for systematic reviews.

1.1.1 Search Strategy

We applied the following search strategies in March 2024. The Dimensions database has been chosen as it offers extensive coverage across a wide range of academic disciplines. It includes not only journal articles but also books, chapters, conference proceedings, and policy documents.

First, we searched for papers concerning the responsible design, development, or use of AI in public sector organizations, which resulted in an initial pool of 1301 papers in total. The keywords covered three aspects of the topic: *AI Technologies, Ethical and Responsible AI, Governance and Public Sector*. Although our selection for terms related to ethics and different principles of RAI may not cover all possible perspectives, we believe that the criteria presented here are widely accepted and could serve as a good starting point. It follows the principles previously proposed in the literature, such as the report on Trustworthy AI by the EU and an overview of AI ethics guidelines by Hagendorff (2020).

Figure 1. Search Terms Used

Databases	*Search terms in the title/keywords*
Dimensions	("artificial intelligence" OR "AI" OR "machine learning" OR "Large language model*" OR "Natural language processing" OR "algorithmic decision-making" OR "algorithmic governance") AND (ethic* OR responsib* OR trustworth* OR accountab* OR privacy OR fair* OR safe* OR transparen* OR robust* OR just* OR explainab* OR "human oversight" OR "human autonomy" OR diversity OR discriminat* OR "bias") AND ("AI governance" OR "Artificial Intelligence Governance" OR "public admin*" OR "public sector" OR "public service*" OR "public agenc*" OR "public organisation")

The following inclusion and exclusion criteria have been used to guide the selection process. The paper should:

1. Explicitly discuss AI systems, not just digital transformations in a general sense.
2. Focus on the process of development, design, or use of AI within an organizational context.
3. Discuss AI within the public sector organizations.
4. Address responsible AI, or mechanisms, approaches for achieving Responsible AI outcomes.
5. Empirical study, by which papers working with primary data are considered.

Based on the abstract screening, 95 papers have been selected. After assessing the relevance and quality of these papers based on full paper screening, 30 papers have been selected for the final study. The most common reason for exclusion was that the studies were not empirical, leading to over 400 papers being excluded.

2 - Preliminary Findings

Our preliminary findings are primarily descriptive, concerning to RQ3. Our analysis of initial set of papers indicates a significant and sustained increase in publications from 2015 to 2023. We have also observed that out of 30 papers, only 5 have used theoretical frameworks to guide their research. It suggests a disconnect between conceptual and empirical studies, which could be explained by lack of maturity of the field.

Additionally, that most papers have been excluded because they are not empirical studies or do not discuss development, design, or use processes. This indicates a lack of focus on practical, actionable insights in the literature. The significant lack of empirical studies can also be attributed to challenges in accessing data from public organizations.

The distribution of papers illustrated that Responsible AI governance research is concentrated in developed regions with strong academic and technological infrastructures, such as North America and Western Europe. Comparing these to the geographic distribution of the final selected papers, it could be argued that countries with high publication density but fewer case studies (e.g., China) might focus more on theoretical research rather than specific governance-related case studies. However, it should also be noted that, the foreign language papers have been excluded in the selection process.

We have also observed that terms like 'transparency' and 'accountability' frequently co-occurred with 'algorithmic decision-making,' indicating automated decision making seen as the major focus within public sector context.

For the final outcome of the research we aim for context-mechanism-outcome framework based on the comprehensive analysis of the papers through a realist synthesis review lens. We choose to have bottom up and iterative approach, to get most out of empirical data from the papers rather than taking pre-defined conceptual analysis model. However, the work is ongoing, we briefly describe some aspects of the contexts and mechanisms we have observed as part of the preliminary results.

For the context, we have identified three major layers: technical (AI lifecycle), organizational and systematic.

In the context of the AI lifecycle, our preliminary findings reveal a significant disconnect between the design and implementation phases. There is a gap between the principles and values aspired to in the design process of AI systems and the outcomes of implementation, where complex real-life scenarios bring unexpected uses and challenges for the technology (Fest, 2023). Additionally, post-production checks have been given the least attention in the finally selected papers, however One successful example emphasized how post-production checks can lead to re-designs of the system to then quickly fix the shortcomings.

Organizational contexts involve aspects such as stakeholders, organizational readiness and culture. Our preliminary results suggest that roles and responsibilities regarding various aspects of responsible AI are unclear, leading to confusion among stakeholders. This often results in data scientists and technical personnel being solely responsible for the outcomes. Based on semi-structured interviews, which are the most commonly used methodology in empirical papers, we have observed that mostly technology and data specialists have been interviewed. Although managers are occasionally included in these studies, they face the challenge of navigating the landscape with less technical understanding of the product and unclear expectations from different stakeholders.

Preliminary results also indicate that the absence of organizational mechanisms, such as knowledge sharing and process standardization, can lead to varied behaviors among street-level bureaucrats, and undermining public values of fairness.
Within a systematic context, we have grouped things happening outside the organization, such as legislation or the political and social landscape. While the legislative context is much clearer, an example of the political context could include the pressure to innovate. Otherwise, the organization would be left behind in practical terms, such as not getting funds from central governments. Public scrutiny could also be considered at the macro level. We plan to group the mechanisms in the same way. The next step involves connecting mechanisms derived from the analysis of the papers to the contexts, allowing us to explain where these mechanisms work or don't work and their outcomes based on the selected empirical works. Additionally, we plan to search different databases to check for any relevant grey literature that may have been missed.

Author Contribution: Conceptualization: A.G., H.v.v.f., N.G., A.V.; data curation, formal data investigation and analysis A.G.; Methodology: A.G., N.G., writing – original draft: A.G.; writing- review & editing: A.G.; H.v.v.f., N.G., A.V.

References

Bertot, J. C., Jaeger, P. T., & Hansen, D. (2016). The impact of policies on government social media usage: Issues, challenges, and recommendations. Government Information Quarterly, 29(1), 30-40. https://doi.org/10.1016/j.giq.2011.04.004
Dignum, V.: Responsible Artifcial Intelligence: How to Develop and Use AI in a Responsible Way. Springer, New York (2019)

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1
Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30, 99–120. https://doi.org/10.1007/s11023-020-09517-8
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2
Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1, 501–507. https://doi.org/10.1038/s42256-019-0114-4
Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Science and Engineering Ethics, 26(4), 2141-2168. https://doi.org/10.1007/s11948-019-00165-5
Pawson, R. (2005). Evidence-based policy: A realist perspective. Sage Publications.
Wong, G. (2013). Realist methods in medical education research: What are they and what can they contribute? Medical Education, 47(1), 89-96. https://doi.org/10.1111/medu.12094
Zhu, W.: 4 Steps to Developing Responsible AI. World Economic Forum (2019). https://www.weforum.org/agenda/2019/06/4-steps -to-developing-responsible-ai/.
Zuiderwijk, A., Chen, Y. C., & Salem, F. (2023). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, 40(1), 101-111.

Fair AI for All: Gender Equity and Socio-Cultural Factors in Sub-Saharan Africa.

Lilian Olivia **Orero**
SafeOnline Women Kenya (SOW-Kenya)

Sub. No: 3577 -St4

Abstract

Introduction
Artificial Intelligence (AI) is progressively playing a vital part for African governments; but, to promote AI decisions as fair, efficient, and proper, becomes challenging. This research targets the Sub-Saharan Africa region because this area has a high population density and a diverse cultural zone, and many of the countries in this region are considered developing countries. The objective of the research is to examine the socio- cultural factors that define advancement of gender equity concerning the use of AI in decision making tools in this setting. In its pursuit of understanding the interplay between AI and governance this study aims to explain the factors that can create or hinder women's involvement in decision-making processes regarding AI technologies in a bid to enhance gender mainstreaming in AI advanced technologies.
Research/Policy Question
The central question of this research is: How do socio-cultural factors in Sub-Saharan Africa influence the promotion of fair AI access and gender equity? Specifically, the study interrogates cultural expectations, social attitudes, and economic factors within which AI is envisioned and employed. This examination is important to understand how they influence the adoption and effect of AI technology in governance, especially gender equity. These concerns also extend to analyzing the efficiency of AI in mitigating gender biases in governance. This entails assessing how these technologies can alleviate current existing disparities in decision making processes and devise ways of advocating for equal representation. In addition, the paper reveals the challenges that eliminate women from participating in AI technologies creation and regulation as well as the potentialities, which may contribute to their engagement.
Research Methodology and Data Used
This research utilizes a pragmatic, mixed-methods research design that incorporates both qualitative and quantitative methods for data collection of grassroots communities in Sub-Saharan Africa. Part of data collection is structured interviews with women from different communities, the goal of which is to identify personal experiences and issues regarding the utilization of AI solutions and levels of AI literacy; focus group discussions with community members aimed at hearing multiple perspectives on the role of AI in gender equity. Complementing this, quantitative methods involve distributing surveys to a larger population sample to collect data on education levels, economic status, and attitudes towards AI technology. The data gathered is subjected to thorough statistical analysis in order to find connections and trends between socio-cultural factors and women

involvement in AI. This comprehensive method combines different approaches to create a strong foundation for grasping the intricate relationship between socio- cultural factors and gender equity in the governance of AI.

Literature Review and Key Theories

The methodology for the study is grounded on key theories in gender studies and technology adoption including the Technology Acceptance Model (TAM) (Davis, F. D. 1989) and the Social Role Theory (Eagly, A. H., & Wood, W.2012). TAM explains how users come to accept and use technology, emphasizing perceived ease of use and usefulness as primary factors influencing adoption. In Sub-Saharan Africa, cultural and educational contexts are highly interwoven with factors relating to AI. Social Role Theory suggests that cultural norms and expectations shape societal roles and explains how women's involvement in AI is influenced by gender roles. Prior studies have demonstrated that inclusive design practices and gender-sensitive policies can help reduce gender biases in AI (Neupane, B., & Sibal, P. 2021).

Related Work

A large number of scholars have conducted research proposing the ways in which gender and technology are intertwined in developing countries. UNESCO's report "Cracking the Code: Precisely due to this collection of articles entitled “Girls’ and Women’s Education in STEM “presents the difficulties and opportunities women face regarding STEM education on the global level. Furthermore, the World Bank has also initiated the study on ‘Digital Dividends’ whereby the role of emerging technologies such as AI can bring enhanced development by solving issues to do with gender equity. These studies give a baseline knowledge of the challenges and potential solutions for gender equity in AI decision making algorithms.

Key Findings

The study highlights key observations on the cultural context influencing gender equity in AI across Sub-Saharan Africa. Women's involvement in AI development is impeded by societal and cultural obstacles like traditional gender norms and a lack of opportunities for STEM education. Financial limitations also hinder advancement, worsening the existing gender gaps. Nevertheless, opportunities and effective programs exist, such as educational initiatives and community involvement, fostering inclusivity and empowering women in the AI sphere. Supportive policies are pivotal, offering financial aid and training programs. Additionally, AI holds promise in mitigating gender biases in decision-making, contingent upon inclusive design and contextual considerations.

Recommendations

The study recommends various key measures to promote gender equity in AI governance in Sub-Saharan Africa. To guarantee inclusivity and cultural sensitivity, it is crucial to promote community involvement by actively including local groups, particularly women, in the development of AI. Moreover, it is imperative to bolster socio-economic expansion by putting into practice focused policies and initiatives that address financial barriers and provide possibilities for women to pursue education and training in the AI industry. Finally, it is crucial to promote social and cultural changes, pushing for shifts in society views and norms to allow women to take on leadership and decision-making positions in the AI landscape.

Conclusion

In conclusion, this study emphasizes the significance of thorough policies and measures that back social, cultural, and economic shifts to improve gender equity in AI governance in Sub-Saharan Africa. By overcoming the obstacles that have been identified and advocating for inclusive methods, Africa can create AI technologies that are just and advantageous for all its residents. Continued initiatives are necessary to guarantee that AI systems have a positive impact on gender equity, promoting a future that is more inclusive and fairer. The research emphasizes the significant role of community engagement, economic growth, and cultural changes in reaching these objectives.

References

1. Davis, F. D. (1989). Technology acceptance model: TAM. Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption, 205, 219.
2. Eagly, A. H., & Wood, W. (2012). Social role theory. Handbook of theories of social psychology, 2, 458-476.
3. Neupane, B., & Sibal, P. (2021). Artificial intelligence needs assessment survey in Africa. UNESCO Publishing.
4. UNESCO. (2017). Cracking the Code: Girls' and Women's Education in STEM. UNESCO Publishing.
5. World Bank. (2016). Digital Dividends. World Development Report.

Gaming Data: Digital urban twins, open data platforms and the ethics surrounding governing data

Fran **Meissner**[1], Florence **Chee**[2] and Michael **Nagenborg**[1]

1. *University of Twente*
2. *Loyola University Chicago*

Sub. No: 1586 -St4

Abstract

As the smart city hype is ebbing, many cities across the globe are more data rich than ever. Despite disillusionment with smart city imaginaries, cities have increasingly been fitted with a plethora of sensing technologies. Many cities have their own open-data platforms that, with reasonable reliability, generate an abundance of real-time data ranging from the geolocation of public transport and people to how air quality and temperature is changing. Even cities with few in-situ sensors can be sensed from the sky with Earth Observation imagery increasing in accuracy. Arguably in response to this availability of data and a continued drive to improve analytics – mostly through machine learning applications - cities are one of the frontiers that require our attention in developing critical global perspectives on the governance of data and AI. Our paper develops the idea that thinking about cities works to encourage a multifaceted approach when thinking about how AI systems are changing governance processes through data.

Urban digital twins are increasingly pushed as a set of technologies that can be used both to monitor cities in real time and to support urban planning. One step removed from ideas about smart cities, where technology would heal all ailments of the urban eradicating the need for governance, the digital twin of a city is meant to allow for both the virtual destruction of the city and its real improvement. The ability to test the impact of interventions – those planned and those that are unplanned like natural disasters – are meant to improve urban development. Mayors and other city officials will draw on insights from twins in order to make decisions about the actually existing city.

In the case of urban digital twins, data - and thereby the city - is made to govern through novel interfaces. This is increasingly done via games engines: a suite of tools originally designed and used to render gaming content and now used for the visualization of data. The use of off-the-shelf technologies is surprising given the complex nature of cities and the value-laden nature of urban simulations.

Some key debates we discuss here bridge common disciplinary gaps in understanding how digital twins shape our collective, epistemic knowledge about cities. As run-ins between embodied realities and serious games/simulations in VR/XR have shown, the foundational models used in everyday civic applications may be built upon assumptions that have the potential to elicit knee-jerk reactions and hasty policymaking. The reason for our concern specifically lies with the power asymmetries that emerge as a result of the entanglements of urban data with AI and the rendering of its output. We will focus on what we can learn from the critical study of games for getting at some of the core questions of new entanglements between data, models and how cities make decisions. For example, in line with Galloway (2006), we argue that game engines are designed to create an illusion of 'continuity' rather than highlighting differences in the quality and quantity of data. Decisionmakers may be disproportionately 'persuaded' by data perceived as complete and value free, since the rendering of urban data qua game engines makes it less likely to challenge the layering of data, modelled indexes, machine learning models (most prominently reinforcement learning) and other techniques used to show optimized ways of planning cities. Looking towards public facing twins it further is relevant to consider notions like the digital sublime (Mosco, 2004) and the sense of magic these new forms of rendering entail. As such there are processes of normalizing large scale data incorporation into the urban experience that can be learned from in deepening our understanding of AI data sets and how to govern them.

Specifically, we wish to highlight how urban institutions of governance are both potentially reshaped and uniquely placed to call for the legal compliance and ethical use of data and AI. The ethics of games can thus teach us about not trying to govern monolithically understood entities like 'the data' or 'the AI' but complex entanglements of data and AI within globally differentiated contexts.

Galloway, Alexander R. (2007) Gaming: Essays on Algorithmic Culture. University of Minnesota Press.

Mosco, V. (2005). The Digital Sublime: Myth, Power, and Cyberspace. MIT.

Political misuses of biometric systems and the (re)production of power asymmetries

Júlia **García-Puig**
Leiden University

Sub.No: 1897 -St4

Abstract

Biometric systems - based on technologies that measure, analyze and process unique biological traits - are transforming the ability of the state to identify individuals and verify their identity. These systems are praised for their reliability, accuracy and security, yet their ongoing expansion raises serious concerns in connection with human rights, civil liberties and social justice (Lyon, 2008; Marciano, 2019; Rao & Nair, 2019). Much of the academic attention has been on the (anticipated) outcomes derived from biometric systems, that is, on explaining *why* these can be dangerous (Marciano, 2019; Strauß, 2023; Williams, 2020), yet less attention has been paid to understanding *how* and *when* biometric systems can grow into abusive and oppressive tools. This paper addresses the latter questions by studying the processes whereby biometric systems turn into instruments of political abuse.
A core issue linked to these systems is that biometric data is a highly sensitive type of personal data (Vacca, 2007). Biometric systems are based on the so-called static bodily characteristics that have unique identifiers and cannot be (easily) modified, such as fingerprints, face, iris pattern and DNA. Biometric systems offer two main functions. One is the verification or authentication of an identity, which consists of verifying that a person is indeed who they claim to be. This is based on a 1:1 (one-to-one) matching, meaning that one's biometric data is only compared to data about themselves that had been previously stored in the system. The second functionality, identification, is a more complex one as it seeks to find out who that person is (Hu, 2017). This relies on a 1:N (one-to-many) matching process, in which one's biometric traits are compared to those of many other individuals available in large databases.
Governmental applications of biometric systems are increasingly being used for numerous purposes, such as to facilitate passport checks in airports and border controls (Sanchez del Rio et al., 2016), to enhance financial inclusion and access to welfare (Rao & Nair, 2019), to distribute aid to asylum seekers (Ajana, 2013), to reduce fraud in elections (Gelb & Clark, 2013), to identify suspects in criminal investigations (Win et al., 2020), and to deploy smart city applications (Bera et al., 2020). Despite their many benefits and promising opportunities, biometric systems are often depicted as intrusive, and as tools used by governments to discriminate and oppress (Lyon, 2008; Marciano, 2019). Such claims are not unfounded; the proliferation of biometric systems has raised concerns in countries around the world and across the political spectrum (De Hert & Bouchagiar, 2022; Martin & Donovan, 2015; Sung, 2023). Amidst such a backdrop, their deployment is particularly concerning in

contexts where governments have a long record of deliberately curtailing freedom of expression and oppressing their populations (Article 19, 2023; Gonzalez, 2023; Liu, 2023).
The analytical focus of this paper shifts away from normative assessments and the prevailing attention on outcomes (De Hert & Bouchagiar, 2022; Marciano, 2019; Martin & Donovan, 2015; Spektor, 2020), to the decisions and steps whereby these can lead to political abuses. In other words, the paper aims to advance the understanding regarding how biometric systems become tools for political mass surveillance, social control and discrimination. To that end, this article presents a conceptual framework to facilitate a nuanced analysis of the decisions that might influence the workings of biometric systems throughout four stages: i) design and deployment of the technology or system; ii) data collection; iii) data analysis; and iv) decision and application. Conceptually, the paper draws on power theory to examine, on the one hand, the role of preexisting power asymmetries between the state and citizens in enabling or constraining political abuses, and, on the other hand, to capture the resulting effects that biometric systems exert on those power configurations.
The framework is not to be understood as an exhaustive one but rather as a comprehensive overview of technical, administrative and political decisions and risks that might be involved in the deployment of biometric systems. These are informed by existing scholarship mainly in the disciplines of critical data studies, surveillance, and citizenship. Notably, this paper builds on the premise that the deployment and outcomes of such systems are highly context dependent and strongly influenced by the sociopolitical institutions and structures in which they are embedded. With that in mind, the framework seeks to facilitate a structured but adaptable approach to conduct empirical and conceptual research in diverse contexts. To that end, I demonstrate the utility of the framework by applying it to the case of China's Integrated Joint Operations Platform (IJOP) in the region of Xinjiang, which consists of a "system of systems" used to build individual profiles by pooling large amounts of data ranging from location, biometrics, behavior, and religious beliefs (Wang, 2019).
The paper provides a stage-based analysis of the IJOP by examining both technical aspects and the sociopolitical context in which it is embedded. In the first stage, I look at the foundations and characteristics of the design and deployment of the system, such as scope and purpose, target population, and institutional context. The second stage analyzes the types of data involved and its collection process. The third stage focuses on how data is analyzed and integrated with other governmental databases. Finally, in the fourth stage, I examine the resulting decisions and effects of the system. The findings illustrate how political intentions are reflected in all four stages, from the deployment of the system, the technology used, the places in which data is collected, and the choice of classifications and categorizations that are then used to flag individuals for police inspection. Considering the outcomes of the system, this paper concludes that the IJOP is influenced by the state's political agenda focused on deepening social control, especially of ethnic minorities.

References

Ajana, B. (2013). Asylum, identity management and biometric control. *Journal of Refugee Studies*, *26*(4), 576–595. https://doi.org/10.1093/jrs/fet030
Article 19. (2023, August 22). *Iran: Tech-enabled 'Hijab and Chastity' law will further punish*

women. Article19. https://www.article19.org/resources/iran-tech-enabled-hijab-and chastity-law-will-further-punish-women/
Bera, B., Das, A. K., Balzano, W., & Medaglia, C. M. (2020). On the design of biometric based user authentication protocol in smart city environment. *Pattern Recognition Letters*,*138*, 439–446. https://doi.org/10.1016/j.patrec.2020.08.017
De Hert, P., & Bouchagiar, G. (2022). Visual and biometric surveillance in the EU. Saying "no" to mass surveillance practices? *Information Polity*, *27*(2), 193–217. https://doi.org/10.3233/IP-211525
Gelb, A., & Clark, J. (2013). *Identification for Development: The Biometrics Revolution* (315). www.cgdev.org
Gonzalez, B. (2023, December 6). *Facial recognition in Iranian Metro being used as scaretactic to enforce hijab*. BiometricUpdate. https://www.biometricupdate.com/202312/facial-recognition-in-iranian-metro-being used-as-scare-tactic-to-enforce-hijab
Hu, M. (2017). From the National Surveillance State to the Cybersurveillance State. *Annual Review of Law and Social Science*, *13*, 161–180. https://doi.org/10.1146/ANNUREV LAWSOCSCI-110316-113701
Liu, Z. (2023). *How Technology Changes Authoritarian State Surveillance: Evidence from China*.
Lyon, D. (2008). BIOMETRICS, IDENTIFICATION AND SURVEILLANCE. *Bioethics*, *22*(9), 499–508. https://doi.org/10.1111/J.1467-8519.2008.00697.X
Marciano, A. (2019). Reframing biometric surveillance: from a means of inspection to a form of control. *Ethics and Information Technology*, *21*(2), 127–136. https://doi.org/10.1007/S10676-018-9493-1/TABLES/1
Martin, A. K., & Donovan, K. P. (2015). New surveillance technologies and their publics: A case of biometrics. *Public Understanding of Science*, *24*(7), 842–857. https://doi.org/10.1177/0963662513514173
Rao, U., & Nair, V. (2019). Aadhaar: Governing with biometrics. In *South Asia: Journal of South Asia Studies* (Vol. 42, Issue 3, pp. 469–481). Routledge. https://doi.org/10.1080/00856401.2019.1595343
Sanchez del Rio, J., Moctezuma, D., Conde, C., Martin de Diego, I., & Cabello, E. (2016). Automated border control e-gates and facial recognition systems. *Computers and Security*, *62*, 49–72. https://doi.org/10.1016/j.cose.2016.07.001
Spektor, M. (2020). Imagining the Biometric Future: Debates Over National Biometric Identification in Israel. *Science as Culture*, *29*(1), 100–126. https://doi.org/10.1080/09505431.2019.1667969
Strauß, S. (2023). The body as permanent digital identity? Societal and ethical implications of biometrics as mainstream technology. *Tecnoscienza*, *14*(1), 59–76. https://doi.org/10.6092/issn.2038-3460/17611
Sung, M. (2023). Questioning the South Korean Smart Border: A Critique of Surveillance Racism, Biometric Identity, and Anti-Immigration. *Korea Journal*, *63*(1), 180–207. https://doi.org/10.25024/kj.2023.63.1.180
Vacca, J. R. (2007). *Biometric Technologies and Verification Systems*. ELSEVIER.
Wang, M. (2019, May 1). *Interview: China's 'Big Brother' App Unprecedented View into*

Mass Surveillance of Xinjiang's Muslims. Human Rights Watch .
Williams, D. P. (2020). Fitting the description: historical and sociotechnical elements of facial recognition and anti-black surveillance. *Journal of Responsible Innovation*, *7*(S1), 74 83. https://doi.org/10.1080/23299460.2020.1831365. Win, K. N., Li, K., Chen, J., Viger, P. F., & Li, K. (2020). Fingerprint classification and identification algorithms for criminal investigation: A survey. *Future Generation Computer Systems*, *110*, 758–771. https://doi.org/10.1016/j.future

Algorithmic Governance

Human-Machine Collaboration for Enhanced Decision-Making in Governance

Dirk Van **Rooy**
Centre for Responsible AI, University of Antwerp, Belgium

Sub.No: DAP-2023-0183 -St5
Full paper is accepted for publication in the Data & Policy journal.

Abstract

A detailed exploration is presented of the integration of human-machine collaboration in governance and policy decision-making, against the backdrop of increasing reliance on artificial intelligence (AI) and automation. This exploration focuses on the transformative potential of combining human cognitive strengths with machine computational capabilities, particularly emphasizing the varying levels of automation within this collaboration and their interaction with human cognitive biases. Central to the discussion is the concept of dual-process models, namely Type I and II thinking, and how these cognitive processes are influenced by the integration of AI systems in decision-making. An examination of the implications of these biases at different levels of automation is conducted, ranging from systems offering decision support to those operating fully autonomously. Challenges and opportunities presented by human-machine collaboration in governance are reviewed, with a focus on developing strategies to mitigate cognitive biases. Ultimately, a balanced approach to human-machine collaboration in governance is advocated, leveraging the strengths of both humans and machines while consciously addressing their respective limitations. This approach is vital for the development of governance systems that are both technologically advanced and cognitively attuned, leading to more informed and responsible decision-making.

Constituting an AI: Accountability Lessons from an LLM Experiment

Kelsie **Nabben**
Max Weber Fellow, Robert Schuman Centre for Advanced Studies,
European University Institute, Florence, Italy

Sub. No: DAP-2023-0123 (Remote Presentation) -St5
Full paper is accepted for publication in the Data & Policy journal.

Abstract

This study explores the integration of a pre-trained Large Language Model (LLM) with an organisation's Knowledge Management System (KMS) via a chat interface, focusing on the practicalities of establishing and maintaining AI infrastructure, as well as the considerations for responsible governance. The research adopts the concept of 'AI as a constituted system' to emphasise the social, technical, and institutional factors that contribute to AI's governance and accountability. Utilising an ethnographic approach, the paper details the iterative processes of negotiation, decision-making, and reflection among stakeholders as they develop, implement, and manage the AI system. The findings indicate that LLMs can be effectively governed and held accountable to stakeholder interests within specific contexts, when facilitated by clear institutional boundaries that foster innovation while navigating risks related to data privacy and AI misbehaviour. Effectiveness is attributed to distinct policy creation processes to guide AI's operation, clear lines of responsibility, and localised feedback loops to ensure clear accountability for actions taken. This research provides a foundational perspective to better understand algorithmic accountability and governance within organisational contexts. It also envisions a future where AI is not universally scaled, but consists of localised, customised LLMs tailored to stakeholder interests.

A systematic review of regulatory strategies and transparency mandates in AI regulation in Europe, the US, and Canada

Mona **Sloane**[1] and Elena **Wüllhorst**[2]

1. *UVA School of Data Science, University of Virginia, Charlottesville VA, United States*
2. *King's College London, London, UK*.

Sub. No: DAP-2023-0171-St5
Full paper is accepted for publication in the Data & Policy journal.

Abstract
In this paper, we provide a systematic review of existing AI regulations in Europe, the United States, and in Canada. We build on the qualitative analysis of 129 AI regulations (enacted and not enacted) to identify patterns in regulatory strategies and in AI transparency requirements. Based on the analysis of this sample, we suggest that there are three main regulatory strategies for AI: AI-focused overhauls of existing regulation, the introduction of novel AI regulation, and the omnibus approach. We argue that although these types emerge as distinct strategies, their boundaries are porous as the AI regulation landscape is rapidly evolving. We find that across our sample, AI transparency is effectively treated as a central mechanism for meaningful mitigation of potential AI harms. We therefore focus on AI transparency mandates in our analysis and identify six AI transparency patterns: human in the loop, assessments, audits, disclosures, inventories, and red teaming. We contend that this qualitative analysis of AI regulations and AI transparency patterns provides a much needed bridge between the policy discourse on AI, which is all too often bound up in very detailed legal discussions, and applied socio-technical research on AI fairness, accountability, and transparency.

Human oversight of algorithmic decisions: a post-deployment empirical investigation.

Susana **Lavado**[1], Charles **Wan**[2] and Leid **Zejnilovic**[1]

1. Nova School of Business and Economics
2. Rotterdam School of Management, Erasmus University

Sub. No: 4411 -St5

Abstract

Human-in-the-loop (HITL) systems represent collaborative partnerships between algorithmic models and humans, acknowledged as opportunities to not only improve the accuracy of algorithmic systems, but also to make human decision-making more effective (Mosqueira-Rey, et al., 2023), and as an attempt to maintain human agency and accountability (Enarsson, Enqvist & Naarttijärvi, 2022), which may be particularly important in high-stakes decisions critically affecting individuals' lives.

The ability of humans to override algorithmic decision, preventing biased or harmful decisions (Green, 2022), is perceived as a safeguard against violations of the rights of data subjects in recent European Union's regulations. The General Data Protection Regulation states that people shall not be subject to exclusively automated decisions that significantly affect them, and the proposal for an Artificial Intelligence Act mandates human oversight for high-risk systems. However, such regulations may reflect expectations that do not necessarily rest upon the reality in machine-human interactions and foster a false sense of security (Digital Future Society, 2022). We aim to contribute to a discussion about human-computer interaction and the oversight of algorithms with a longitudinal analysis of a real-world machine-human supervision interaction.

Context

Until October 2019, the Portuguese Public Employment Service (PPES) used an algorithmic decision-making system (ADMS) to predict the individual risk of long-term unemployment of individuals registered at the PPES. One of the most salient limitations of that system was that counselors tended to uncritically accept algorithmic suggestions: in 2019, only 1.6% of suggestions of the previous system were override by counselors (Zejnilović et al., 2020).

Then, the PPES piloted and deployed a new ADMS, which was in full use in June 2021. Besides improving the algorithmic models in use, the new system encouraged counselors to adopt a supervising role by removing the need for justifying the overriding of the algorithm and by providing training about the characteristics and limitations of the system.

The current study aims to investigate the effectiveness of such measures to increase humans' effective supervision of algorithmic decisions, and whether that potential effectiveness is sustained through time. While we expected an uptick in the overriding rate immediately after the launch of the new system, we hypothesized that, as time passes, override rates would revert to 2019 levels.

Data and Method

Our dataset contained data between June 2021 and December 2022, monthly counts of the number of decisions 1371 counselors made while working in 53 employment centers in continental Portugal. The decisions were grouped by the risk level suggested automatically and the risk level attributed by the counselor, for a total of 2,918,332 decisions.

Some counselors had decisions registered in more than one employment center, but the majority of their decisions occurred at one center. We selected the center where the counselor made more decisions, discarding other observations. We eliminated months where a given counselors had made less than 10 decisions in their main center. The final dataset contained 1113 counselors and 1,948,400 decisions.

We computed the rate of overridden decisions per counselor, that is, the percentage of decisions where the counselor did not accept the risk level suggested by the algorithmic model.

Results

Figure 1 shows the evolution of the average override rate per month. As hypothesized, in the first month, the rate equaled approximately 29%, a strong uptick from about 2% before the new system was implemented. In the first six months, there was a steady decrease of about 2 percentage points per month. After, it starts to plateau until it reaches around 13% in the last three months.

Discussion

The overriding rate dropped sharply before plateauing six months after the deployment of the system. Two conflicting hypotheses may explain these results. On one hand, it is possible that while the release of a new system may have motivated counselors to critically assess the algorithmic output, that effect diminished after such novelty effect passed. This hypothesis is consistent with qualitative findings at the PPES suggesting that counselors tended to uncritically accept the algorithmic decisions of the previous system (Zejnilovic et al., 2020). On the other hand, counselors may start trusting the ADMS more, accepting more its outputs. Future work may test these hypotheses, exploring whether the reduction in changes improves overall performance of the ADMS. Overriding the algorithm decision is not inherently beneficial, as counselors could reject correct algorithmic decisions. For human supervision to be effective, the overriding of algorithmic decisions needs to be coupled with counselors' capacity to critically assess the algorithmic outputs, disentangling incorrect decisions from correct ones.

Policy makers may be particularly interested in the potentiality of counselors' disengagement from critical evaluation of the algorithmic output. At the research level, this would imply a necessity to explore which strategies prevent that disengagement, such as periodic training sessions, or messages that reinforce the counselors' role as machine supervisors. Furthermore, future research may investigate the circumstances of the algorithmic decisions that typically lead to its acceptance or rejection.

The positive outcome of the results is that, even 19 months after the system was in use, the override rate was still significantly higher than the one observed in the previous system (13% vs. 2%). This result suggests that the removal of barriers to overriding algorithmic decisions and the delivery of training about the limitations of the system were still effective 1.5 years after system deployment.

References

Digital Future Society, 2022. Towards meaningful oversight of automated decision-making systems. https://digitalfuturesociety.com/app/uploads/2022/11/Towards_meaningful_oversight_of_automated_decision_making_systems.pdf

Enarsson, T., Enqvist, L. & Naarttijärvi, M. (2022) Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts. Information & Communications Technology Law, 31:1, 123-153. https://doi.org/10.1080/13600834.2021.1958860

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. Computer Law & Security Review, 45, https://doi.org/10.1016/j.clsr.2022.105681.

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D. et al. (2023). Human-in-the-loop machine learning: a state of the art. Artif Intell Rev 56, 3005–3054. https://doi.org/10.1007/s10462-022-10246-w

Zejnilović, L., Lavado, S., de Rituerto de Troya, Í, Sim, S., Bell, A. (2020). Algorithmic Long-Term Unemployment Risk Assessment in Use: Counselors' Perceptions and Use Practices. Global Perspectives, 1(1). https://doi.org/10.1525/gp.2020.12908

Trust in algorithmic governance: A meta-analysis.

Evrim **Tan**
KU Leuven

Sub. No: 2715 -St5

Abstract

This conference paper presents a comprehensive meta-analysis of trust factors in algorithmic governance, with a specific aim of discerning the impact of policy measures in influencing citizens' trust in algorithmic governance and AI applications. In an increasingly digitalized world, understanding the dynamics of trust in technology-driven public sectors is crucial for enhancing the adoption and effectiveness of algorithmic governance systems. Numerous empirical studies have explored the multifaceted relationship between citizen trust and algorithmic governance, shedding light on the effectiveness of policy measures, the personal characteristics of respondents, and socio-demographic factors. Furthermore, contextual factors and methodological choices have been found relevant to explain this relationship. This meta-analysis synthesizes the findings from these studies and explores the role of independent factors and contextual conditions that can shape the trust-citizen-algorithm relationship.

Theoretically, the paper follows the framework developed by Zucker (1985) and Bodo and Janssen (2022) to categorize policy measures to influence citizen trust under three pillars, namely familiarity, control, and insurance. These categories serve as key constructs for understanding the influence of trust-enhancing policy measures as part of the public policy processes.

Familiarity stands for a shared and stable set of background knowledge and expectations that citizens and public officials have with the algorithmic systems in use. Policy measures that focus on the ex-ante processes in algorithmic governance such as transparency, and explainability of AI or other design-related choices are perceived as policy measures influencing familiarity.

Control pertains to the level of influence that citizens have over the algorithmic processes directly or indirectly through a custodian. Studies have shown that citizens who perceive themselves as having more control over algorithmic decision-making are more likely to place their trust in these systems. This aspect includes policy measures at an interim stage where citizens engage with the algorithm and incorporate policy measures concerning data sharing, human involvement in decision-making, and the use of data-control enhancing technologies such as blockchain.

Insurance, the third pillar, refers to the extent to which individuals are protected against negative outcomes or consequences of algorithmic decisions. Trust in algorithmic governance is positively influenced when citizens feel adequately protected against potential negative consequences. Ex-post measures such as appeal possibility, legal assurance and other regulatory guarantees are categorized in the insurance category.

Under these three pillars of trust, several hypotheses have been tested through meta-analysis and metaregression about the impacts of policy measures (i.e familiarity, control, and insurance) on citizen trust, and the moderating/mediating roles of personal traits, socio-demographic, and contextual factors.

The findings from meta-analysis and meta-regression suggest that all types of policy measures (familiarity, control, insurance) are effective on citizen trust in algorithmic governance. This effect holds for both attitudinal and behavioural trust and appears to be independent of the context. Among the three types of policy measures, insurance appears to be the most influential measure for citizen trust.

Except for perceived benefits, personal traits such as technology knowledge or privacy concerns, do not have a significant influence in explaining the variance in citizen trust. Furthermore, age has been found the only socio-demographic variable that has a limited significant impact on citizen trust if trust is measured as behavioural trust.

The significant predictive power of the perceived benefit of algorithms suggests that wider usage of algorithmic governance and the perceived improvement in public services can increase trust, with or without policy measures. This finding also complies with several theories such as the exposure effect, social learning theory, cognitive dissonance theory, and technology acceptance model. This warrants caution from a public policy perspective because self-governance/laissez-faire options do not appear viable to regulate algorithmic governance. Rather legislative action clarifying the insurance mechanisms against the potential fallacies of algorithmic governance comes as the most viable policy option to increase citizen trust in the long term.

References:

Balázs Bodó, Heleen Janssen, Maintaining trust in a technologized public sector, Policy and Society, Volume 41, Issue 3, September 2022, Pages 414–429, https://doi.org/10.1093/polsoc/puac019

Zucker L. G. (1985). Production of trust: Institutional sources of economic structure, 1840 to 1920. In L. L. Cummings & B. Staw (Eds.), Research in organizational behavior (pp. 1840–1920). JAI Press.

Towards Algorithmic Accountability in the Public Sector.

Simone Maria **Parazzoli**
ISI Foundation

Sub. No: 4001 -St5

Abstract

The pervasive influence of algorithms in our societies has transformed the way we communicate, analyse the world, conduct business, and make decisions across various domains. Governments, recognising the potential benefits, are increasingly embracing algorithmic technologies to enhance administrations' operational efficiency, service delivery, and responsiveness to citizens. This adoption, particularly in automating decision-making processes, holds promise for improved public services while raising concerns about accountability. This research aims to address the central question of how the public sector can integrate algorithms in its practice without compromising accountability.

Taking a policy-oriented perspective, the study delves into the role of algorithms in the public sector, offering an overview of algorithm's role in the public sector. Further, it explores the meaning of algorithmic accountability in the public sector in light of the literature on accountability in public administration. Finally, it provides an original mapping of algorithmic accountability policies based on an innovative automated decision-making system life cycle, bridging the gap between theoretical principles and practical policy solutions.

The first section evaluates the benefits and risks of the use of algorithms in the public sector, analysing their impact through the lens of public value, focusing on efficiency, service delivery quality, trust and legitimacy, and outcome achievement. Risks, particularly associated with transparency, fairness, and accountability, are identified, emphasising the pivotal role of automated decision-making systems as sources of risks within the broader landscape of algorithms in the public sector.

The second section explores the meaning of algorithmic accountability in the public sector, interpreting it as the intersection of public accountability and algorithmic accountability. The principal-agent theory and the social contingency model of public accountability are discussed in relationship to algorithmic accountability. Further, this concept is analysed in light of the five dimensions specified by Bovens as determining the heterogeneity of accountability in practice: actor, forum, their relationship, the details of the account, and its consequences.

The third section presents a comprehensive mapping of algorithmic accountability policies, structured around an original automated decision-making system lifecycle, which was built combining technical and policy lifecycles into a socio-technical structuring of the process. The mapping facilitates a nuanced understanding of policy options available to policymakers, considering their advantages, disadvantages, and impact on various stakeholders. Nineteen algorithmic accountability policies are identified and discussed.

Further, steps towards evidence-based regulatory decisions in the realm of automated decision-making systems are suggested. Given the goal of identifying which algorithmic accountability policies work best, hence how to best set up the regularory constraints to ADMSs' actors behaviour to balance benefits and risks, two crucial questions are identified and discussed. First, what criteria should be adopted to evaluate the success of a combination of policies? Second, what evidence can be used to measure the result of an algorithmic accountability policy or combination of policies on those criteria?

This research contributes to the ongoing discourse on the responsible implementation of automated decision-making systems in the public sector, guiding policymakers toward informed decisions that balance the benefits and risks associated with these transformative technologies.

How to construct a trustworthy AI ethical principle: Inspired by Feenberg.

Xiaomei **Wang**[1] and Huayu **Xin**[1,2]

1. *Zhejiang University*
2. *University of Edinburgh*

Sub. No: 7309 -St5

Abstract

Identifying a mechanism to make artificial intelligence (AI) "trustworthy" poses a significant challenge in the realm of contemporary ethics. The quest involves the creation of a concrete ethical code to address the moral dilemmas emerging from AI's rapid development. Such guidelines must reflect and promote the public interest, yet their formulation often falls prey to biases favoring the interests of dominant groups. This bias leads to what is termed an "iron cage" scenario, where the ethical principles are unduly influenced by these groups, diminishing their credibility and effectiveness.

In response to this, a dynamic balance mechanism, the "technical code," is proposed as a means to regulate multi-party interests. This approach is seen as a vital tool to prevent the entrenchment of fixed interest patterns and, thereby, dismantle the "iron cage" of control. The technical code concept, rooted in the works of philosopher Andrew Feenberg, suggests embedding cultural values and moral considerations into the very fabric of technology. This method not only democratizes the design process of AI but also ensures that technology becomes a carrier of these values.

The concept of trustworthiness in AI is explored in-depth, emphasizing the need for AI systems to be controlled, reliable, comprehensible, and respectful of human rights and democratic values. The potential moral hazards associated with AI, such as the infringement on privacy, personal autonomy, and lifestyle choices, are acknowledged. The study advocates that AI should remain under human oversight to mitigate these risks.

The philosophical underpinnings of trust and technology are examined through the lens of thinkers like Immanuel Kant, Thomas Hobbes, and Karl Marx. Their ideas provide a historical and ethical context to the modern challenge of creating AI ethics. The evolution of society's understanding of trust and ethical codes is traced, highlighting a shift from divine command or universal order towards more pragmatic and interest-based models.

A critical analysis of the social, political, and cultural environments that shape the development of ethical codes is presented. This study identifies the structural limitations within modern societies, where technology and bureaucracy combine to form a restrictive "technical cage." This cage often limits public participation and perpetuates the interests of powerful groups, particularly in the realms of capital and technology. These imbalances significantly influence professional fields and decision-making processes.

Our study concludes by emphasizing the critical role of public participation in the design of AI technology. This participation is essential at both the micro and macro levels to avoid marginalization of certain groups and to ensure the creation of ethical principles that are genuinely trustworthy. The technical code is championed as a means to break free from the restrictive "iron cage" of current technological governance. This approach aims to democratize technology, ensuring that the development of AI not only reflects but also promotes public interests, democratic freedoms, and the common good of society.

Overall, the proposed framework for constructing trustworthy ethical principles for AI centers around balancing diverse interests, integrating ethics from the outset, and fostering broad-based public participation in technology design. This framework addresses the complexities of AI development and offers a path towards ethical principles that are credible, balanced, and representative of a wider array of societal interests.

Global Challenges & Dynamic Threats

AI and Digital Transformation of the Greater China Region: A Comparative Study of AI strategies in China, Taiwan, and Hong Kong.

Wilson **Wong**[1], Charles **Hinnant** [2]and Natalie **Wong**[3]

1. *The Chinese University of Hong Kong*
2. *Florida State University*
3. *National Chengchi University*

Sub. No: 0465 -St6

Abstract

This paper would like to contribute to the theory and practice of artificial intelligence (AI) and the digital transformation of governments by examining the approach and focus in the AI strategies of the three public administration systems of the Greater China Region: China, Taiwan, and Hong Kong. With the rapid advancement of AI technology, there is little doubt that AI will create unprecedented impacts on societies, economies, and governments (Henman, 2020; Lee, 2018; Mergel et al., 2016; Young et al., 2019). Although there is a growing volume of literature on AI and its application in the public sector, few theory-guided and empirical-based explanatory theories are built (Bareis & Katzenbach, 2022; Guenduez & Mettler, 2023; Madan & Ashok, 2022; Zuiderwijk et al., 2021). To fill this gap, this paper would assess the AI strategies of the three systems in practice under the guidance of the public administration and AI governance literature.

As a study integrating the theory of public administration and AI governance with empirical analysis, the basic research question we would like to address can be stated as simple as what changes are being made by government in AI strategies and to what extent they are consistent with what is prescribed in the literature. To go further, the more specific questions are: what are the strategic concerns of the AI development of these systems? What is the role of public administration and its reforms for AI transformations?

In the “Race to AI”, when nations and governments are competing in the AI frontier, they are eager and active in turning their mission, vision, and plans into strategies that are widely available online on their websites for establishing their legitimacy and mandate (DiMaggio & Powell, 1983; Galindo et al., 2021; Smuha, 2021). The availability of important information in the form of documents or websites about AI strategies by governments has enabled researchers to measure the existence of any gap between what governments are doing and what the literature has suggested they should. This would also facilitate the process of building more solid and reliable theories in guiding governments for having the foresight to master the technology.

With the approach and focus of AI development as the key concern in our study, two particular issues deserving our special attention are social equity and digital transformation of government as they are the core issues in Public Administration as well as the studies of AI policy and governance. The attainment of social equity is always a core mission of all governments. At the same time, AI is taken as a disruptive technology because it is anticipated to transform the structure and organizations of public administration to significantly enhance its performance (Gritsenko & Wood, 2022). In this connection, the study would like to assess the level of attention and preparation of governments in addressing issues of the digital divide and the transformation of public organizations in the AI era.

Although there has been an increasing amount of research on the impact of AI, especially on the technical end and economic growth and innovation, there are still some major gaps and unanswered questions. Research on AI from the social science perspective and the vantage point of public administration in particular is still rare and in a burgeoning stage. Besides, to our knowledge, no study has focused on the Greater China Region. Under these circumstances, some common limitations are frequently observed in reviewing the current state of the study of AI strategies and policies of countries. First, many of them are descriptive surveys to identify, summarize, and categorize the major AI initiatives (Bareis & Katzenbach, 2022; Galindo & Sheeka, 2021). Despite being able to provide a good account of national AI policies and development on an international level or even a global scale, many of them are not yet at the advanced stage of knowledge discovery and generation with theory-building and hypothesis testing.

When the above type of research tends to be more on the empirical side, some theory-oriented studies discuss the expected impacts of digital transformations of organizations under the impact of AI from a more normative standpoint (Wong & Chu, 2020). These two types of studies are often not well-connected. This creates a normative-empirical gap, meaning that there is little research that integrates both theory and empirical analysis in a single study. Even when there is a theory-based empirical analysis, they are often studies based on a generic approach of organizations or experiences of the private sector (Khanal, 2022; Madean & Ashok, 2022; Wirtz et al., 2019), not on the core knowledge of public administration which recognizes the main pillars of the field such as the uniquenesses of government and public-private differences.

In the research design, the selection of the three public administration systems in the Greater China Region represents a strike of balance between heterogeneity and homogeneity in terms of contexts in comparative studies. Furthermore, the comparative study design enables us to find out if there are any variations among similar with also distinctive systems. Although China, Taiwan, and Hong Kong all belong to the Greater China Region, they have their own public administration systems with individual traditions and special features. It permits the testing of the main hypothesis of the technology enactment framework (Fountain, 2004) that divergence in applications of information and communication technologies (ICTs) is possible among public administration systems due to the presence (or absence) of institutional factors which vary across institutions and contexts (Janowski, 2015; Wong & Hinnant, 2022).

Through the analysis of AI strategies, new light can be shed on developing a more comprehensive and empirically-based theoretical framework to identify the challenges for public administration in the AI era. Well-researched case studies of nations and governments can produce theoretical insights for scholars and actionable lessons for policymakers. The study and its discernible findings should facilitate the effort of taking advantage of the theories and knowledge of Public Administration and AI governance in building more robust and useful theories to guide the development of AI in and by governments.

China's dynamic data infrastructuring process: genealogy of blockchain hype and how it's intertwined with today's AI development and governance.

Zichen **Hu**
London School of Economics and Political Science

Sub. No: 0629 -St6

Abstract

Summary: This paper starts by analysing blockchain hype in China for two reasons. First, decentralisation, free-market, autonomous, and liberal imaginaries that feature the hype seem contradictory to China's (PRC's) high degree of political centralisation. Second, there is a mismatch between the public (technology journalism and individual investors) and the official discourse on and prospect of blockchain. These two significant yet perplexing mismatches are shown in China's in-flux data infrastructure, with components of blockchain—and then AI and many others—added due to the need assumed to represent the 'maximised' 'public good'. This leads to an ongoing process of data 'infrastructuring' (Hartong & Piattoeva, 2021). This process involves various stakeholders who propose different normative values and regulatory approaches to disruptive technologies through which social action and social order can take place.

A discontinuous and erratic process of negotiation is sustained in the multistakeholder discourse. The shifting governance paradigm of blockchain technologies in China forms through an interplay between (1) contestations on normative concepts related to 'decentralisation' (e.g., privacy, autonomy, civil rights, power distribution structure, etc.), and (2) political economic (conflict of) interests related to the dynamics of fictitious capital (DeFi) and new ways of social organisations (DAOs). The shifting governance paradigm of blockchain, negotiated between state and non-state actors, between the decentralising agent/potential of emerging technologies and centralising political power, can provide critical lessons for understanding and addressing the ongoing challenges associated with the AI surge and its governance.

Conceptual Framework and research questions:

Issues of public value 'break out' around the emergence of blockchain technologies because they (1) afford new types of political and financial participation and (2) symbolise new patterns and paradigms of imaginaries of privacy, autonomy, power distribution, and social order. In this process, different actors produce discourse on blockchain technologies that represent different understandings of 'public good' and deliberate choices with particular social, cultural, and political consequences. These include increasingly prominent voices, both from the private and public sectors in China, that advocate integrating blockchain technologies with AI in both market-oriented applications and the e-government system. However, the tension is evident between the private and public sectors

in terms of different objectives, normative values, and political economic interests, which might sustain today's AI hype with implications for power dynamics.

Thereby, my paper asks:

RQ1: To what extent do blockchain and decentralised platforms afford the formation of a public and symbolise new patterns and paradigms of imaginaries of privacy, autonomy, power distribution, and social order?

RQ2: How does the constant struggle amongst multiple stakeholders between the decentralising agent and centralising political power sustain today's AI hype and governance in China?

Methodology:

My approach is to map discursive regimes (inspired by Bauer & Schiele, 2023), based on the understanding of discourse as a field of strategy, a space of struggle and contestation, meaning that it can be used for competing purposes or combined in incompatible ways (Foucault, 1981, 1994). By'mapping discursive regime', I mean to investigate the multistakeholder discourse of blockchain technologies as an assemblage infused with specific power relations that encompass both discursive and non-discursive elements, bearing within themselves traces of past practices and discourses (see Potts, 2019, pp. 91-108).

Specifically, the scope of the discursive regime is based on a body of documents (now 50), including state-level government policy documents, prominent and high-reputational tech journalism (where tech startups make their voices), and third-party consulting and law firms' analysis reports. It is noteworthy that the data repository is still expanding, rendering the research an ongoing project.

The multistakeholder perspective entails contesting accounts of the justificatory logics of governing paradigms and imaginaries around privacy, security, fairness, and potential risks, which are reconfigured by decentralised network infrastructure. This also adds to the geopolitical scrutiny of the China-US technology rivalry by introducing the complex domestic power dynamics and competing ideologies inferred from the multistakeholder discourse analysis. Analysing these discursive interactions and intertextuality implicated in discursive networks does not just help trace to 'origin stories', but also translate the economic and political conditions of their production. The next step will be analysing the congruence or conflict networks at the topical level and longitudinal versions of these networks with visualisation (Leifeld, 2017).

Contribution:

My research emphasises how industries and regulatory authorities bargain and negotiate the room for development and application of emerging and potentially disruptive technologies like blockchain and the ethical dimensions of different ways of addressing 'public good' as part of the justificatory logic. To do so, I identify the discontinuous and erratic processes of emerging technologies negotiated in multistakeholder discourse. Such processes underpin the ad hoc and compromised ways that China's blockchain data infrastructure has developed, which highlights a number of underlying sociotechnical conditions that lead to an ongoing process of 'data infrastructuring' (Hartong & Piattoeva, 2021). For example, this dynamic process involves advocates for integrating blockchain into AI applications and AI-aided governance. Such advocates are contextualised in the ongoing

challenges associated with the AI surge and its governance, especially the constant struggle between the decentralising agent and centralising political power amongst multiple stakeholders. The state's normative values and political economic interests favouring power concentralisation, demonstrated in the not-so-successful blockchain hype in China, all make the future of the seemingly 'new' and promising AI a bit unsure, despite the fact that China is now the second biggest source of AI talent and research output. One of the challenges is that to commercialise AI against the extremely high cost of training LLMs, situational adaptation of the model and interoperability amongst the models are needed and can be realised through blockchain technologies. Will large models be financially sustainable, and will China's AI future be one with autonomy? These questions remain lingering. Therefore, instead of idealising China's AI hype, it is more fruitful to situate it in the genealogies of the discursive boundary work on the potential and risks of emerging technologies, delineate the complex interaction of transformations, and therefore map their possible opportunities and challenges in the future.

The 'Coup' Season: What Can Machine Learning and Artificial Intelligence Tell Us About the Resurgence of Coup D'Etats in Africa?

Elikplimi Komla **Agbloyor**[1], Boakye **Danquah**[1], Agyapomaa Gyeke **Dako**[1] and Lei **Pan**[2]

1. *University of Ghana Business School*
2. *Curtin University*

Sub. No: 7406 -St6

Abstract

Despite the global trend towards increased democracy, Africa has witnessed a resurgence of military coups in recent years, raising concerns about peace, democratic stability, and economic progress on the continent. Since 1960s, the African continent has experienced 214 out of 487 global attempted or successful military coups, impacting democratic transitions and national stability. While the 1990s and early 2000s saw a decline in coups due to democratic transitions and strengthening institutions, the trend has reversed with 11 coup attempts recorded in 2020-2023. In this study, we aim to predict the propensity of a country to experience a coup, exploring baseline risk factors and short-term triggers. Utilizing machine learning techniques capable of modeling thousands of variables, we investigate the incidence of coup attempts in the region taking cognizance of country and region-specific differences. We used data spanning 1960 to 2022 for the 54 sovereign African countries. Our baseline estimate finds that coup d'état occurrences are instigated by several key factors, including a historical pattern of coups, a country's experience with democracy, the prevalence of political terrorism, and structural crises like inequality, economic recession, unemployment, and inflation. We also observed that variables like agricultural value-added and mineral rents play a role in fomenting coup attempts within the region. Additionally, our findings indicate that colonial origin serves as a fundamental factor in explaining the occurrence of coups in the region.

Resilience of territories in the face of hydrogeological risk: the role of mitigation interventions in Lombardy region

Giovanni **Azzone**[1], Teresa **Bortolotti**[2], Giulia **Piantoni**[1], Sara **Ratti**[1], Piercesare **Secchi**[2]

1. *Department of Management, Economics and Industrial Engineering, Politecnico di Milano (Italy)*
2. *MOX -Department of Mathematics, Politecnico di Milano (Italy)*

Sub. No: 0432 -St6

Abstract

1. **Introduction**

Climate change is increasingly affecting territories (Westerhoff et al., 2021), enhancing, a.o., hydrogeological risk (HR; Ellena et al., 2020). As other countries, Italy is facing an escalation of these phenomena and their consequences (Gariano and Guzzetti, 2016): HR is high in the 19% of its territory and exhibits an increasing trend (Triglia, et al., 2021). Moreover, socio-economic and demographic complexities (e.g., a densely populated territory with intricate structural characteristics) intensify the effects of HR and their potential relation to the systemic resilience (SR) of municipalities (Tian and Lan, 2023). Indeed, the literature often indicates HR as a potential determinant of resilience of a territory (Beccari, 2016; Azzone, et al., 2022). However, no structural evidence of the relationship between HR and other determinants of SR (of municipalities) is available.

Thus, we formulate as first question: *is HR related to SR of a municipality? How?*

Furthermore, we know that HR can be mitigated by specific interventions (mitigation interventions), but, by now, there is no evidence on their potential effects on the SR of a territory. Thus, our second question is: *to what extent do interventions of HR mitigation affect SR of a municipality?*

2. **Methodology**

Our analysis focuses on the Lombardy region in Italy, relevant as it is experiencing an escalation in hydrogeological phenomena[1] (Gariano and Guzzetti, 2016) and given its socio-economic significance: our statistical units are the municipalities of Lombardy in 2022 (N=1.506).

In line with our aims, we constructed a dataset comprising: descriptive variables on general characteristics of municipalities (*source: ISTAT*), variables on the hydrogeological risk (*source: ISPRA*) and on publicly financed interventions for mitigation (*source: Rendis*), determinants of resilience (Table 1). As there is no unique definition of resilience (Martin and Sunley, 2015), we identified its determinants and created a synthetic indicator, as the direct sum of the standardized determinants per municipality[2].

Table 1. Determinants of resilience

DIMENSION	Determinant	Indicator	Source
SOCIAL	Structural Dependence	Ratio of population of non–working age (0–14 and 65 or older) to population of working age (15–64), times 100	*Istat, 2020*
	Employability	Number of employees per municipality	*Istat, 2019*
	Integration	Proportion of foreign residents	*Elaborations on data from Istat, 2023*
	Healthcare services	Distance to the closest hospital	*Elaborations on data from Ministry of health, 2019*
	Social services	Per capita resources to close the social services gap	*MEF, 2019*
	Competences	High level education: % of college enrolees	*Elaborations on data from Istat, 2017*
ENVIRONMENTAL	Land-use	Building expansion index in population centres and cores	*ISPRA, 2022 (increase 21-22)*
INFRASTRUCTURE	Access	Access to transport networks	*Elaborations on Istat, 2022*
	Digital	Landline internet coverage greater than 100 Mbps, per household residents (FTTH and FTTC)	*Elaborations on AgCom Data, 2019*
ECONOMIC	Financial resilience	Share of turnover at risk in local enterprises	*Elaborations on AIDA, 2022*
	Competitiveness	Indicator of specialization in medium–high–technology or knowledge–intensive sectors	OpenData Regione Lombardia, 2019
	Income	Taxable income per taxpayer	*MEF, 2021*

Then, we verified and quantified the relationship between resilience of a municipality and its exposure to HR. Grounding on the classes of risk identified by Arena et al. (2023), we employed the Mann-Whitney U test to assess the mean difference in terms of resilience and its determinants across the different risk classes.

To quantify the impact of mitigation interventions on resilience, we considered only the public funds allocated between 1999 and 2022 for completed interventions and we constructed: (i) a dichotomous variable (1 if the municipality received at least one funding), and (ii) a continuous variable determined as the weighted average of the funding received by the municipality itself and its neighboring municipalities (Gaussian kernel)[3]. A simple linear regression model for resilience is then fitted, considering either the public funding or one of its transformations as the independent variable.

3. Findings

All determinants of resilience, but the share of turnover at risk and the proportion of foreign residents, are significantly different at a 1% significance level between municipalities with medium-low risk and those with high and top risk. Specifically, municipalities with high and top-high risk are associated with determinants that adversely affect resilience, compared to medium-low risk municipalities. Also, the test on the mean difference in resilience reveals with high significance that resilience is lower in high and top-risk municipalities compared to medium-low risk municipalities (p-value = 0, mean difference is 4)[5]. Last, it emerges that higher funding is significantly associated with a reduction in resilience and with an increase in the social and material vulnerability, as quantified by Didkovskyi et al. (2020).

4. **Discussion Conclusion**

We found a relation between HR and resilience of municipalities, thus confirming the risk classes in Arena, et al., (2023) and our synthetic indicator of resilience. Also, it emerges that mitigation interventions cannot enhance – in the short term – resilience: being related to HR, mitigation interventions result allocated to municipalities with high risk and low resilience. Therefore, for enhancing SR (decorrelating it and HR), other interventions are needed, mainly on structural dependence, accessibility and specialization.

This work provides an original definition of resilience of municipalities in the face of disruptions and risks and a composite indicator for assessing it. Also, the dataset we constructed bolsters a quantitative approach on a regional scale at a municipality level, representing a novelty in the reference literature.

Also, policy makers are provided with insights on SR, its relationship with HR and the role of public interventions, shaping the discourse on resilience of municipalities facing natural disasters.

References

Azzone,G., De Felice,G., Pammolli,F. (2022). *Data for Italy: Rapporto sull'Economia dei Territori*: Il Mulino.

Arena,M., Azzone,G., Bortolotti,T., Piantoni,G., Ratti,S., Scotti,F., Secchi,P. (2023). *Dissesto idrogeologico in Lombardia. Studi e ricerche sui temi prioritari del Programma Regionale di Sviluppo Sostenibile*. Polis Lombardia.

Beccari,B. (2016). A comparative analysis of disaster risk, vulnerability and resilience composite indicators. *PLoS currents*,*8*.

Didkovskyi,O., Azzone,G., Menafoglio, A., Secchi, P. (2021). *Social and Material Vulnerability in the Face of Seismic Hazard: An Analysis of the Italian Case,* Journal of the Royal Statistical Society Series A: Statistics in Society,184(4),1549–1577.

Ellena, M., Ricciardi, G., Barbato, G., Buffa, A., Villani, V., Mercogliano, P. (2020). Past and future hydrogeological risk assessment under climate change conditions over urban settlements and infrastructure systems: the case of a sub-regional area of Piedmont, Italy, *Natural Hazards,* 102,275-305.

Gariano, S.L., Guzzetti, F. (2016). Landslides in a changing climate, *Earth-Science Reviews*,162, 227-252.

Martin, R., and Sunley, P. (2015), On the notion of regional economic resilience: conceptualization and explanation, *Journal of Economic Geography,* 15:1–42.

Tian, N. and Lan. H. (2023). The indispensable role of resilience in rational landslide risk management for social sustainability, *Geography and Sustainability,* 4,70-83.

TrigliaA., Iadanza C., Lastoria B., Bussettini M., Barbano A. (2021) *Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio* - Edizione 2021. ISPRA, Rapporti 356/2021.

Westerhoff , L., Carina E., Keskitalo, H., and Juhola, S. (2011) Capacities across scales: local to national adaptation policy in four European countries, *Climate Policy*, 11(4),1071-1085.

Future-Proofing Data Governance to Prepare for Climate Change.

Jacob **Leiken**[1], Beverley **Hatcher-Mbu**[2] and Tom **Orrell**[2]

1. *NYU School of Law*
2. *Development Gateway: An IREX Venture*

Sub. No: 0860 -St6

Abstract

Policy Question

To fight climate change effectively, actors must deeply understand the nature of emerging, and often unpredictable, changes to the planet. Gaining this understanding in order to pilot and scale solutions is only possible with data. Devising effective policy responses to mitigate the worst effects of climate change and inform adaptation strategies will require agriculture data, water data, land management data, early warning system data, and more. These data will be needed at the local, national, subregional, continental, and global levels. The data will also need to be provided in a standardized, quality-ensuring format with well-defined indicators. It is essential that as much data as possible is made accessible, as ingesting more kinds of data improves predictions about weather, water levels, and crop yields.

Considering the broad scope of these data, climate change may present the biggest data challenge ever faced. And it is paired with an equally great data governance challenge: "interoperability" is increasingly at the forefront of the data conversation, but progress in this area is moving too slowly. In the words of one minister of Artificial Intelligence: "For one day on Earth, there is around 100 petabytes of data created... It would take a scientist 100 years to analyse one day's worth of climate and meteorological data. Climate change is a race against time and if we do not have the systems that are able to read this data, crunch it and give us advice on a real-time basis, we are losing the race."

How can we start now to future proof data governance structures in preparation (and immediate need) for the growth of climate data?

Methodology

To understand the data transparency and governance landscape, we looked at development projects related to climate change in Africa from a number of other organizations, including carbon markets, agriculture tools, and more. We were particularly interested in the data transparency of each project. The catalog created by this research is accessible here.

We drew additional learnings while working on projects in the climate sector. The Great Green Wall Accelerator Multipurpose Platform, in collaboration with the UNCCD, Pan African Agency for the Great Green Wall, and the eleven Great Green Wall nations, allows

donors, partner states, and implementing agencies to share data and information. We are also co-developing a data governance framework to support new and existing digital livestock data platforms in close collaboration with Ethiopia's Ministry of Agriculture via the aLIVE program.

Key Findings

Through our work on the above projects, we came to the following conclusion: Climate data is the newest category, and AI is the newest application, but data use has a long history that we can build from when we're thinking about effective governance needed to power AI in climate decision making going forward. There is so much we can build from. Acting on this conclusion can take a number of forms, explored below.

Governments should move away from individual data sharing, when possible, towards institution-wide or cross-institutional data sharing. The investment required to securely implement access and privacy protocols across whole governments is seen as one of the biggest barriers to making accessible troves of climate data a reality. The aLIVE program, a partnership between Ethiopia's Ministry of Agriculture and Development Gateway, aims to create a livestock information system which provides timely, relevant, and interoperable data on livestock. We have begun the process of developing context-specific data-sharing protocols, and will share in real-time our lessons learned from building buy in and broad consensus to support data sharing in practice.

Nonprofit organizations, such as Development Gateway, should support governments to take "whole of ministry" approaches in developing, scaling, and connecting various digital agriculture tools. For example, ministry-wide and national strategies need to lay the foundation for the growth of data lakes, which will be a critical data source powering effective AI in identifying climate needs. Development Gateway is working with Tanzania on their first digital agriculture strategy, in which we encourage an emphasis on cross-sectoral infrastructure investments.

Developers must use existing systems, instead of building new systems from scratch. Although trite, it is imperative that we do not waste time reinventing the wheel. There are many sector-specific working groups, data steward structures, and so on, in which data governance decision-making is decided upon collectively across institutions and needs. These efforts can be scaled into structures that better reflect the diversity of climate data needs. The Global Partnership for Sustainable Development Data's Effective and Ethical Data Sharing at Scale Cookbook provides a cross-sector approach to establishing collaborative and working group structures. It also includes several examples from the US and globally of those structures and how they approach key issues around effective data.

Countries with data protection policies based on the EU's GDPR may need additional legislation encouraging open data and data transactions. The EU has recently enacted both the Data Governance Act and Data Act, which aim to facilitate safe data-sharing in various sectors. The AI Act, currently in trilogue, will impose additional regulations on AI to ensure its safe development and usage. Not all three acts need to be, or should be, enacted in jurisdictions with GDPR-inspired data protection laws; however, clear, concise policy or guidance around how to comply with local data protection law is essential to promote data-sharing.

Our presentation will discuss each of these approaches in-depth, and how Development Gateway is orienting our work to promote clear and consistent data governance to power AI using climate data.

An integrated decision support tool for assessing the risk of labour exploitation on fishing vessels.

Ruoyun **Hui**[1], Jamie **Hancock**[1], Jat **Singh**[1,2], Hannah **Thinyane**[3], Mark **Briers**[1] and Anjali **Mazumder**[1]

1. Alan Turing Institute
2. *University of Cambridge*
3. *Diginex*

Sub.No: 5990 -St6

Abstract

Introduction

Labour exploitation and related human rights abuses in the fishing industry has received increasing attention in the past decade, after appalling cases being brought to light by the media and NGOs [1]–[6]. Closely related to other illegal activities at sea (e.g., drug and human smuggling), it causes severe harm to the human dignity of individual victims, the marine ecosystem, the livelihood of coastal communities, and national security.

The low transparency of working and living conditions on isolated fishing vessels at sea poses a major challenge in tackling labour exploitation at sea. Many suspicious fishing activities take place on the high seas where monitoring and enforcement effort is weak. The complex regulatory and judicial landscape formed by the flag state, coastal state, port state and regional fisheries management organisations (RFMOs) adds to the difficulties in coordination, as does the extensive interconnections between labour exploitation and economic development, marine ecosystems, migration, and law enforcement internationally.

Recent years have seen much development around using data to better elucidate fishing activities, in particular AIS data, which include time-stamped location information of the vessels. There exist publicly available data sources on subjects such as fishing effort [7], exposure and response to IUU fishing by states [8], and the risk of human rights abuses in seafood supply chains [9]. Other studies have leveraged data analytics to look into flags of convenience [10], [11], transshipments [12], [13], hidden activities on the high seas [14] and near Marine Protected Areas [15], [16], and IUU fishing and labour abuse [17]–[19]. Commercial tools such as Ocean Mind [20] and Starboard [21] are also available to provide more detailed intelligence service to analysts. However, large-scale risk assessment by authorities like port state control is still mostly based on limited static information. The data-

driven, technology-laden response to human rights abuses has also been criticised for marginalising situated knowledge and streamlining surveillance [22].

This paper describes the development of a transparent and highly customisable open-source decision support tool that leverages AIS and other data sources without losing sight of contextual knowledge and human expertise, to aid state authorities and civil society in the combat against labour exploitation at sea.

Materials and methods

Through a literature review, we identified five main clusters of concerns that shapes the risk of labour exploitation on fishing vessels: vessel validity, owner and crew, working and living conditions, fishing activities, and history and connections. These are used as top-level indicators in the decision support tool. Each is then elaborated by a number of secondary indicators, which further develop into operational-level indicators that directly maps onto data or information sources investigators might access, including AIS data, shipping registry data, state-level indicators, IUU fishing lists and direct observations. Although unstructured and not routinely available, we also incorporated potential information from open-source investigations such as social media search and witness testimony into the model.

We used a Bayesian network to represent the dependencies and hierarchical structures among the indicators. It can effectively cope with missing data and flexibly accommodate arbitrary functional relationships between levels of indicators. We produced a working version of the tool by plugging in the data sources and functional relationships of our choice, although we expect investigators to critically review and adapt it according to the context they work in. We would like to facilitate such collaborative exercises in our future work.

To streamline customisation, we also developed a web-based dashboard for displaying the risk model and input data. The hierarchical model structure can be edited in plain text. The weights and scores of each indicator can all be updated from the web interface, which also generates an intuitive representation of the sources of risks.

Results

We tested the working version of the decision support tool on positive (labour exploitation reported), negative (not engaged in labour exploitation), and unknown cases. We show that the decision support tool is able to capture suspicious cruising patterns in the positive cases, although data paucity in most other sources poses a challenge.

Mindful that the result might reflect a focus on distant water fleet fishing on the high seas due to heightened media attention, we also illustrated how an investigator aware of a newly introduced visa loophole not reflected in the built-in data sources might adjust the model to apply to the working conditions of migrant fishing workers in Ireland.

Discussion

The decision support tool based on a Bayesian network can assist large-scale risk profiling for prioritising inspections or further investigation. Compared to commercial services, it is cost-effective, transparent, highly flexible, and retains the agency of interpretation with the investigator. The same framework can be potentially adapted to other domains of human rights investigation too.

Developing and testing the decision support tool highlight that despite the promise brought out by detailed location data, other crucial aspects of the fishing industry necessary to detect labour exploitation on fishing vessels remain data-poor. For example, vessel registry

data are rarely publicly available; even where they are, tracing the beneficiary ownership of fishing vessels can be extremely difficult due to complex ownership structures and bank secrecy laws. Transparency in seafood supply chains also remains notoriously low. Effective fisheries governance requires policy changes and international cooperation to promote information sharing, as some countries have taken the initiatives to do.

Related to the incomplete sources of data, the decision support tool is primarily designed for risk profiling instead of building evidence for individual cases. Direct evidence for exploitation - such as working conditions and employment relationships - are rarely obtainable without onboard investigation. Nevertheless, by including indicators in the model that are contingent on open sources or inspections, we aim to promote a more consistent investigation workflow.

The landscape around labour exploitation at sea is rapidly shifting as in other adversarial scenarios. The actors, behaviours, policies and data sources are all changing dynamically according to the local context. Therefore, we consider it of paramount importance to engage closely with practitioners and those most impacted to build tools they can adapt and apply most effectively.

References

[1] Environmental Justice Foundation, 'Sold to the Sea: Human trafficking in Thailand's fishing industry', 2013. [Online]. Available: https://ejfoundation.org/resources/downloads/Sold_to_the_Sea_report_lo-res-v2.compressed-2.compressed.pdf

[2] Environmental Justice Foundation, 'Blood and Water: Human rights abuse in the global seafood industry', 2019. [Online]. Available: https://ejfoundation.org/resources/downloads/Blood-water-06-2019-final.pdf

[3] E. Htusan and M. Mason, 'More than 2,000 enslaved fishermen rescued in 6 months', *Associated Press*, Ambon, Indonesia, Sep. 17, 2015. Accessed: Dec. 06, 2022. [Online]. Available: https://www.ap.org/explore/seafood-from-slaves/more-than-2,000-enslaved-fishermen-rescued-in-6-months.htm

[4] Issara Institute and International Justice Mission, 'Not in the Same Boat: Prevalence & Patterns of Labor Abuse Across Thailand's Diverse Fishing Industry', 2017. [Online]. Available: https://globalinitiative.net/analysis/not-in-the-same-boat-prevalence-patterns-of-labor-abuse-across-thailands-diverse-fishing-industry/

[5] Greenpeace, 'Misery at sea: human suffering in Taiwan's distant water fishing fleets', 2018. [Online]. Available: https://www.greenpeace.org/aotearoa/publication/misery-at-sea/

[6] Greenpeace, 'Seabound: The journey to modern slavery on the high seas', 2019. [Online]. Available: https://www.greenpeace.org/static/planet4-southeastasia-stateless/2019/12/c4f6f6b4-greenpeace-seabound-b.pdf

[7] D. A. Kroodsma *et al.*, 'Tracking the global footprint of fisheries', *Science*, vol. 359, no. 6378, pp. 904–908, Feb. 2018, doi: 10.1126/science.aao5646.

[8] G. Macfadyen, G. Hosch, N. Kaysser, and L. Tagziria, *The IUU fishing index, 2019. Poseidon Aquatic resource management limited and the global initiative against transnational organized crime*. Global Initiative. https://globalinitiative. net/wp-content/uploads/2019/02 ..., 2019.

[9] B. Krogh-Poulsen, S. L. McDonald, and T. Woodcock, 'Seafood Social Risk Tool: Identifying risk of forced labor, human trafficking, and hazardous child labor in the seafood industry', Monterey Bay Aquarium, Jul. 2022. [Online]. Available: https://www.seafoodwatch.org/our-projects/seafood-social-risk-tool
[10] J. H. Ford and C. Wilcox, 'Shedding light on the dark side of maritime trade – A new approach for identifying countries as flags of convenience', *Marine Policy*, vol. 99, pp. 298–303, Jan. 2019, doi: 10.1016/j.marpol.2018.10.026.
[11] G. A. Petrossian, M. Sosnowski, D. Miller, and D. Rouzbahani, 'Flags for sale: An empirical assessment of flag of convenience desirability to foreign vessels', *Marine Policy*, vol. 116, p. 103937, Jun. 2020, doi: 10.1016/j.marpol.2020.103937.
[12] N. A. Miller, A. Roan, T. Hochberg, J. Amos, and D. A. Kroodsma, 'Identifying Global Patterns of Transshipment Behavior', *Front. Mar. Sci.*, vol. 5, p. 240, Jul. 2018, doi: 10.3389/fmars.2018.00240.
[13] Global Fishing Watch, 'Revealing the Supply Chain at Sea: A Global Analysis of Transshipment and Bunker Vessels', Apr. 2021. [Online]. Available: https://globalfishingwatch.org/wp-content/uploads/Global-Transshipment-Analysis-Reveals-the-Supply-Chain-at-Sea.pdf
[14] T. H. Frawley *et al.*, 'Clustering of disaggregated fisheries data reveals functional longline fleets across the Pacific', *One Earth*, vol. 5, no. 9, pp. 1002–1018, Sep. 2022, doi: 10.1016/j.oneear.2022.08.006.
[15] S. McClatchie, 'Distant-water industrial fishing in high diversity regions', Oceanography, preprint, May 2021. doi: 10.1002/essoar.10507004.1.
[16] T. D. White *et al.*, 'Tracking the response of industrial fishing fleets to large marine protected areas in the Pacific Ocean', *Conservation Biology*, vol. 34, no. 6, pp. 1571–1578, Dec. 2020, doi: 10.1111/cobi.13584.
[17] G. G. McDonald *et al.*, 'Satellites can reveal global extent of forced labor in the world's fishing fleet', *Proc. Natl. Acad. Sci. U.S.A.*, vol. 118, no. 3, p. e2016238117, Jan. 2021, doi: 10.1073/pnas.2016238117.
[18] E. R. Selig *et al.*, 'Revealing global risks of labor abuse and illegal, unreported, and unregulated fishing', *Nat Commun*, vol. 13, no. 1, p. 1612, Dec. 2022, doi: 10.1038/s41467-022-28916-2.
[19] J. Park *et al.*, 'Tracking elusive and shifting identities of the global fishing fleet', *Sci. Adv.*, vol. 9, no. 3, p. eabp8200, Jan. 2023, doi: 10.1126/sciadv.abp8200.
[20] 'OceanMind', OceanMind. Accessed: Nov. 27, 2023. [Online]. Available: https://www.oceanmind.global
[21] 'Starboard Maritime Intelligence'. Accessed: Nov. 27, 2023. [Online]. Available: https://starboard.nz/#
[22] M. M. Bennett, J. K. Chen, L. F. Alvarez León, and C. J. Gleason, 'The politics of pixels: A review and agenda for critical remote sensing', *Progress in Human Geography*, vol. 46, no. 3, pp. 729–752, Jun. 2022, doi: 10.1177/03091325221074691.

Understanding Discrepancies Between Self-reported and Measured Climate Shocks in Small-scale Agriculture.

Didier **Alia**[1], C. Leigh **Anderson**[1], Joaquin **Mayorga**[1], Rebecca **Toole**[1], Andrew **Tomes**[1] and Stanley **Wood**[1]

1. *University of Washington*

Sub. No: 8944 -St6

Abstract

Introduction

Agricultural producers face increasing risk from the impacts of climate variability. Small-scale producers (SSPs) engaged in rain-fed subsistence agriculture, as is common in sub-Saharan Africa (SSA), are particularly exposed and vulnerable to major livelihood and food supply losses (McCullough, 2017; Dosio, 2017; Azzarri and Signorelli, 2020). Strategies to manage these risks are possible at both the individual and policy level, the former relying on the willingness and ability to engage in adaptive behaviors, and the latter relying on understanding SSP behavior change in response to changes in institutional incentives, such as national social insurance policies. For both, the starting point is associating anticipated climate shocks to SSP decision-making, expected to be a function of past climate experiences (Amare and Balana, 2023). But initial analyses find a mismatch between self-reported climate shocks and measured weather data. Understanding the basis of the discrepancy is vital for making optimal policy decisions in the face of a changing climate and its impact on agricultural communities.

The discrepancies may indicate a data measurement error from a spatial or temporal mismatch between the granularity of the measured data and the household data. In this case, investments in local weather stations or other more finely calibrated data would improve our ability to predict and thereby prepare for adverse events within microclimates. Or the differences may originate at the respondent level, driven by recall error or decision biases such as anchoring or reference dependence (Guiteras et al., 2015). Studies of farmers across multiple countries suggest variability in climate change perceptions (Li et al. 2013; Ogalleh, et al., 2013; Below et al., 2012) driven by individual or household characteristics including gender, age, income, weather information and extension contact (Mengistu, 2011; Deressa et al., 2011 and Apata, 2011). Further, when shocks are under-reported, they could also signify successful adaptation strategies. Understanding whether this variation is individually idiosyncratic, or if there are patterns in SSP driven variation, is central to constructing effective policies.

Our research seeks to contribute to this nascent literature by explaining the discrepancy between self-reported and satellite or sensor measured climate shocks in SSA teasing out what is attributable to data sourcing differences and what is due to variation in how individuals experience weather (Guiteras et al., 2015; Cullen and Anderson, 2017; Nguyen and Nguyen, 2020). We focus on precipitation and temperature extremes, that significantly impact agricultural systems (Funk et. al, 2022). We expect this work to be directly policy relevant, responding to the "call for more local-level analyses to gain a better understanding of the fundamental processes underlying adaptation and for better targeting of adaptation policies by national and local governments, NGOs and bi-lateral donors" (Below et al. p. 224 and citing Boko et al., 2007; Mano and Nhemachena, 2007; Smit and Wandel, 2006). We additionally fill a gap by examining the differences in adaptation behaviors when using self-reported versus measured climate shocks; analyzing how such shocks influence intercropping, improved seed adoption, fertilizer use, planted area, and labor decisions. Understanding individual SSP variation informs the requisite scale for risk diversification at the policy level, and the returns to data investments relative to other policy interventions.

Data and methodology

We combine high-quality household surveys with satellite weather imagery. We use household-level survey data from the World Bank's Living Standard Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) surveys for Ethiopia, Nigeria, Tanzania, Uganda, and Malawi from 2009 to 2019. The panel surveys are nationally representative, geo-referenced, and consist of detailed data on socioeconomic characteristics, farming indicators, and self-reported weather shocks. For measuring climate shocks, we rely on satellite weather data from the Climate Hazards Infrared Precipitation with Stations (CHIRPS; Funk et al., 2015). We use daily rainfall measurements at a resolution of 0.05-degrees to calculate indicators of exposure to climate shocks, following Anderson et al. (2023), Nguyen and Nguyen (2020), and Verdin et al. (2015) for data-driven definitions of climatic extreme exposure.

Our empirical approach involves two main equations. Our first panel estimation models self-reported shocks against measured shocks and controls:

y_{it}^1=\alpha_0+\alpha_1\mathrm{shock}_{it}^1+\alpha_2{\mathrm{share}}_{it}+\alpha_
+\gamma\ X_{it}+\varepsilon_{it} (1) (See PDF for a readable version of each formula)

Where y_{it}^1 is a dummy equal to one if the household i self-reported a shock in period t, {shock}_{it}^1 is a dummy equal to one if the household was exposed to a measured shock in period, {share}_{it} is the share of years in the most recent 20 years when the household experienced a shock, \alpha_i is a household-specific intercept, X_{it} is a vector of time-varying household controls, and \varepsilon_{it} is an i.i.d. error. We use our second equation to examine how perception or exposure to shocks is related to farm- or household-level adaptations:

y_{it}^2=\beta_0+\beta_1\mathrm{shock}_{it}^2+\beta_2\mathrm{share}_it+\beta_i+\delta\
mathrm {Z}_{it}+\eta_{it} (2)

Where y_{it}^2 is a variable representing adaptation, such as an indicator of using intercropping, the log of planted area, the log of on- or off-farm labor, and others . On the right-hand side, \beta_i is a household-specific intercept, Z_{it} is a vector of household-level controls and \eta_{it} is an i.i.d. error. We estimate two variations of this equation, one

where {shock}_{it}^2 represents self-reported shocks and a second one where that variable represents measured shocks. We then compare the coefficients from the estimations with different definitions of the {shock}_{it}^2 variable.

Potential for generating discussion

We expect to find a positive but less than perfect relationship between self-reported and measured shocks across all countries and waves, but the source of the discrepancy is likely to vary by region and shock. For example, frequency may lower the likelihood of report, as habitual exposure may prompt adaptations that mitigate adverse impacts and reduce shock salience, but as recent work in Malawi has found, repeated and broadly co-variant exposure can also deplete a household's ability to adapt or rely on neighbors, resulting in unsustainable coping strategies such as reducing consumption (McCarthy, 2020).

Our study can provide evidence to stimulate discussions on the role of investing in more accurate climate data to shape agricultural policies for SSP resilience relative to the more tailored support in risk communication to manage different perceptions of weather events.

Anticipating Migration for Policymaking

Where Forecasting and Foresight Meet Data and Innovation: Toward a Taxonomy of Anticipatory Methods for Migration Policy

Sara **Marcucci**[1] and Stefaan **Verhulst**[1]

1. *The Governance Lab, New York, United States of America*

Sub: No: DAP-2023-0173 – Sp2
Full paper is accepted for publication in the Data & Policy journal.

Abstract

The various global refugee and migration events of the last few years underscore the need for advancing anticipatory strategies in migration policy. The struggle to manage large inflows (or outflows) highlight the demand for proactive measures based upon a sense of the future. Anticipatory methods, ranging from predictive models to foresight techniques, emerge as valuable tools for policymakers. These methods, now bolstered by advancements in technology and leveraging non-traditional data sources, can offer a pathway to develop more precise, responsive, and forward-thinking policies.

This paper seeks to map out the rapidly evolving domain of anticipatory methods in the realm of migration policy, capturing the trend towards integrating quantitative and qualitative methodologies and harnessing novel tools and data. It introduces a new taxonomy designed to organize these methods into three core categories: Experience-based, Exploration-based, and Expertise-based. This classification aims to guide policymakers in selecting the most suitable methods for specific contexts or questions, thereby enhancing migration policies.

Augmentation or Replication? Assessing Big Data's Role in Migration Studies

Tuba **Bircan**
Department of Sociology, Vrije Universiteit Brussel, Brussels, Belgium

Sub. No: DAP-2023-0088 – Sp2
Full paper is accepted for publication in the Data & Policy journal.

Abstract

As the field of migration studies evolves in the digital age, big data analytics emerge as a potential game-changer, promising unprecedented granularity, timeliness, and dynamism in understanding migration patterns. However, the epistemic value added by this data explosion remains an open question. This paper critically appraises the claim, investigating the extent to which big data augments, rather than merely replicates, traditional data insights in migration studies. Through a rigorous literature review of empirical research, complemented by a conceptual analysis, we aim to map out the methodological shifts and intellectual advancements brought forth by big data. The potential scientific impact of this study extends into the heart of the discipline, providing critical illumination on the actual knowledge contribution of big data to migration studies. This, in turn, delivers a clarified roadmap for navigating the intersections of data science, migration research, and policymaking.

Developing AI predictive migration tools to enhance humanitarian support. The case of EUMigraTool

Cristina Blasi **Casagran**[1], Mr Georgios **Stavropoulos**[2]

1. Autonomous University of Barcelona, Barcelona, Spain
2. Information Technologies Institute Centre for Research and Technology Hellas, Thessaloniki, Greece

Sub. No: DAP-2023-0095– Sp2
Full paper is accepted for publication in the Data & Policy journal.

Abstract

The EUMigraTool (EMT) provides short-term and mid-term predictions of asylum seekers arriving in the European Union, drawing on multiple sources of public information and with a focus on human rights. After three years of development, it has been tested in real environments by 17 NGOs working with migrants in Spain, Italy and Greece.

This paper will first describe the functionalities, models, and features of the EMT. It will then analyse the main challenges and limitations of developing a tool for non-profit organisations, focusing on issues such as (1) the validation process and accuracy and (2) the main ethical concerns, including the challenging exploitation plan when the main target group are NGOs.
The overall purpose of this paper is to share the results and lessons learned from the creation of the EMT, and to reflect on the main elements that need to be considered when developing a predictive tool for assisting NGOs in the field of migration.

Mobile phone data for anticipating displacements: Practices, opportunities, and challenges

Bilgeçağ **Aydoğdu**[1], Özge **Bilgili**[2], Suphi **Güneş**[3], Albert Ali **Salah**[1]

1. *Computing and Information Sciences, Utrecht University, Utrecht, Netherlands*
2. *Interdisciplinary Social Science, Utrecht University, Utrecht, Netherlands*
3. *Turkcell Technology, Istanbul, Turkey*

Sub. No: DAP-2023-0084– Sp2
Full paper is accepted for publication in the Data & Policy journal.

Abstract

The question of how big data sources can address knowledge gaps in migration studies has not been fully answered. While there is increasing research on how anonymised and aggregated mobile phone data (MPD) can be used to develop migration and mobility indicators, a comprehensive understanding of the potential opportunities and challenges of using MPD in displacement research is lacking. In this paper, we review the process of acquiring, processing and analyzing MPD sources for displacement.We present how MPD can serve as a tool for anticipatory analysis in response to natural disasters and conflicts that cause internal or cross-border displacement. Next, we turn our attention to the challenges of using MPD for policy-making, in particular with regards to user privacy and data ethics.We conclude by discussing the potential avenues for future research, bolstered by first-hand experiences in past and ongoing data collaboratives with telcos.

Mixed-frequency VAR: A new approach to forecasting migration in Europe using macroeconomic data

Emily **Barker**[1] and Jakub **Bijak**[1]

1. *The University of Southampton*

Sub. No: DAP-2023-0108– Sp2

Abstract

Forecasting international migration is a challenge that, despite its political and policy salience, has had limited success so far. In this paper we take an approach that employs a range of macroeconomic data to represent different drivers factors of migration. We also take into account the relatively-consistent set of migration policies within the European Common Market, with its constituent freedom of movement of labour. Using panel vector autoregressive (VAR) models for mixed-frequency data, we forecast migration in the short- and long-term horizons for 26 of the 32 countries within the Common Market. We also demonstrate how the methodology can be used to assessing the possible responses of other macroeconomic variables to unforeseen migration events -- and vice versa. Our results indicate reasonable in-sample performance of migration forecasts, especially in the short term, although with varying levels of accuracy. They also underline the need for taking country-specific factors into account when constructing forecasting models, with different variables being important in various regions of Europe. For the longer term, the proposed methods, despite high prediction errors can, be still useful as tools for setting coherent migration scenarios.

Could we have seen it coming? Towards an early warning system for asylum applications in the EU

Emily **Barker**[1] and Jakub **Bijak**[1]

1. *The University of Southampton*

Sub. No: 9304 – Sp2

Abstract

Forecasting large changes in the numbers of asylum applications, an element of so-called asylum `crises', is challenging. Attempts to employ early warning systems date back at least to the large inflow of asylum seekers into Europe in 2015--16, which was relatively unforeseen. In this paper, we present a model that shows that the warning signs of a crisis could appear in publicly-available macroeconomic, geopolitical, and demographic data sources, including some `big data' collections. We propose and test an early warning system for asylum applications in the EU that would be easy to use, effective and interoperable for policy makers, and that would give sufficient advance warning that authorities can be prepared for an increase in the number of asylum applications. We examine whether a model can give a warning signal up to six months in advance for two of the most prominent asylum flows from the recent decade, involving people fleeing the wars in Syria and in Ukraine.

Methodology: In an early warning model with a binary response variable, for each period in the observation window, the binary variable takes a value of 0 to indicate no crisis or 1 for a crisis. The model thus estimates a probability, that a crisis will occur, which will trigger an early warning if this probability is greater than some threshold value calculated specifically for that model. The model calculates three threshold values, one method provides a significantly higher threshold. For each case study, we evaluate its ability to successfully predict the binary response variable of interest. Based on the literature on signal detection, and juxtaposing possible outcomes against prediction in a so-called confusion matrix used for assessing the performance of each model. These measures for evaluating are used in our accuracy analysis through the receiver operating characteristic (ROC) curve and the area under the curve (AUC).

Data: This paper proposes and tests an early warning model for asylum applications with a six-month lead time. The binary indicator is constructed from four measures, all must be met to identify a crisis. The asylum applications must have a current period growth rate over 3, 6, 9 or 12 months of G\%, a growth rate exceeding a value of J\% in the previous 12 months, a number of applications exceeding a of a rolling 12 month standard deviation, and a set minimum value. The crisis indicators all use asylum data at time t. While the explanatory variables are taken from six months previously, so that e.g. a crisis in July 2015 is predicted using the explanatory variables up until January 2015. For each variable (where appropriate),

we include the lag and difference for up to 12 months in the LASSO estimation which selects the explanatory variables. In terms of possible predictors, attempts to forecast civil unrest have been recently made, in particular with reference to the Arab Spring, using data from social media, and (protest) events using the GDELT [Global Database of Events, Language and Tone] data. Further data sources include: Google searches from Google Trends; Ukrainian inflation from the State Statistics Service; US international trade data from FRED St Louis; and exchange rates from the IMF.

Key findings: The models produced promising results, and all give a fairly high degree of predictability, far greater than an even chance as shown in the ROC-AUC analysis. To evaluate the performance of the models, contemporaneous models are presented - there are three contemporaneous models for Syria, and two for Ukraine. There are 3 models that use the explanatory data at a six-month horizon. The models for Syria generally performed better across all measures than the ones for Ukraine, however, accuracies all exceeded 80\%, and all but one for each country, the AUC exceeded 0.9. A possible explanation for the relative better performance for Syrian models, is that the political situation and troubles are available in domestic data only, while Ukraine has a large external 'threat' of Russia. The model analysis presented in this paper has demonstrated that no single model can be useful in every context, with different variables being preferable for different applications, situations and countries. While macroeconomic data might not be the first choice for an array of scholars, there are also important insights that can be learned from them. Importantly, a crucial element of building an EWS model involves desk research on the causes of each of the crises, to identify background and context to find why, and how, these conflicts occurred and escalated. Of course, with hindsight, it is always easy to say that there may have been clear signals at the time. Fully grasping the magnitude of the crisis was not something that could be reliably explained by the model alone. In such applications, only human input would be ultimately able to fully confirm the seriousness of the challenge, whilst remaining cognisant of all the ethical and legal aspects involved in relying on models for helping shape the political or humanitarian responses to the crisis of displacement.

Safeguarding migrant rights through open digital ecosystem (ODE) principles: a prerequisite for deploying anticipatory methods.

Rohan **Pai**[1] and Amrita **Nanda**[1]
1. *Aapti Institute*

Sub. No: 7821– Sp2

Abstract

A growing interest in tracking and managing cross-border migration has pushed nation-states worldwide to introduce digital technologies within their border and migration management processes. At the same time, the predominance of smartphones amongst people on the move has allowed displaced communities to interact online with those who can provide reliable information about safe routes for travel and other necessary services. Through these varied uses of technology by people on the move, we are starting to witness the critical role that digital technologies can play in mediating movement. However, the roles and interests of stakeholders within the migration ecosystem deeply influence technology design, deployment and usage—exacerbating the vulnerabilities of affected populations and people on the move.

States, humanitarian service providers, private sector actors and civil society organisations (CSOs) are critical nodes in enabling human mobility. They actively participate in the migration ecosystem by providing or using digital technologies to track migrants, refugees and asylum seekers. Data assemblages generated through migrants' and refugees' interactions with these digital technologies are leveraged by stakeholders for purposes they see fit. For instance, governments are heavily invested in developing tools for migration forecasting to aid in policy development, resource allocation and emergency preparedness. However, overarching questions persist around the usage of traditional and non-traditional datasets in anticipatory methods, as there is a lack of discourse on the legal and regulatory frameworks that are required to promote migrant rights through safe data collection, analysis, and utilisation. In light of that, this paper aims to explore how large-scale deployment of digital technologies in migration management needs to be preceded by intentional technological design and ethical data governance, to create cross-cutting societal value from emerging migration data infrastructures.

The paper will draw from Aapti Institute's ongoing work on data infrastructures for mixed migration, reflecting on emerging insights through primary and secondary research across three case studies: the forced displacement of Rohingya refugees from Myanmar to Bangladesh; the economic migration of Nepalese workers to India; and Indian students traversing to Germany for higher education. By being anchored in these case studies, the

paper will examine the untapped potential of data that can be harnessed through design principles of open digital ecosystems (ODEs) to enable diverse actors to
a. innovate anticipatory tools to forecast migration movements while safeguarding the rights of people on the move,
b. mitigate harms in the implementation and usage of technologies by relaying agency to migrant communities.
Currently, digital technologies that interact with migrants and refugees are prone to centralised, opaque, and top-down management of data. Human Rights Watch highlights how biometric data of ethnic Rohingya refugees in Bangladesh was collected for identification purposes by the UNHCR and shared without informed consent to state authorities in Myanmar to further repatriation efforts. Similarly, BBC reported how foreign students were unfairly deported from the UK after being falsely accused of cheating in standardised English language tests, based on evidence from voice recognition technologies used to detect proxy test takers. Numerous instances expose the vulnerability of people on the move who are subject to decisions made by prevailing stakeholders through unchallenged access and control over their data. However, our on-ground observations of humanitarian service provision in Rohingya refugee camps within Cox's Bazar also depict the vast potential of digital technologies in streamlining crisis response and forecasting future needs, owing to the capability of these digital technologies to be used as data infrastructures for decision-making.
To enable the ethical use of data infrastructures, technological design principles characteristic of ODEs need to be deployed by multiple actors to effectively use, share and analyse datasets produced by them. These principles include:
a. Openness and Transparency: shifting toward open standards to enable interoperability across digital technologies and allow easier access to datasets across actors in a digital environment
b. Modularity: a disaggregated software stack rather than monolithic architecture, to decentralise and allow multiple stakeholders to participate on a single digital network
By deploying these ODE principles to digital technologies within the migration ecosystem, traditional and non-traditional data sets can be used to unlock societal value and enable community participation. Traditional data, such as government statistics, surveys, census and historical records are publicly accessible information that are being led by the Open Government Data (OGD) movement—a set of policies that promote transparency, accountability and value creation. A study by IOM recommends the use of diverse administrative databases to create comprehensive migration-related statistics. This process envisions the extraction of individual data from different administrative databases to produce reliable figures and describe trends.
Similarly, non-traditional data sources, such as satellite imagery, social media, and mobile application data, can be repurposed towards creating information ecosystems that are led by migrants through the support of agencies on-ground. For example, Humanitarian OpenStreetMap Team (HOT), helps create micro-level maps through crowdsourced data collected on mobile phones, made openly available online for navigation, and shared with community centres, host country governments, and NGOs. Initiatives like these can help

safeguard community interests and build protection around data usage through licensing specifications.

For both, traditional and non-traditional data sources, an interoperable data-sharing ecosystem can be created to enable coordination between entities through open standards. Additionally, if designed using a modular architecture, a digital consent manager for data exchange can be plugged into various points in the digital ecosystem to grant migrants and refugees agency over their data.

An initial scoping has identified how ODE principles shape an entry point into greater negotiating power for people on the move, and help to create an appropriate environment for the ethical usage of their data. This paper will emphasise the need for a principle-first approach at the intersection of mixed migration and digital technologies, by evaluating its effectiveness in unlocking community agency in migration management.

AI, Ethics and Policy Governance in Africa

Are Certain African Ethical Values at Risk from Artificial Intelligence?

S.T. **Segun**
Global Center on AI Governance and The African Observatory on Responsible AI, South Africa

Sub. No: DAP-2023-0153 – Sp3
Full paper is accepted for publication in the Data & Policy journal.

Abstract

This paper questions how the drive towards introducing artificial intelligence (AI) in all facets of life might endanger certain African ethical values. It argues in the affirmative that indeed two primary values that are prized in nearly all versions of sub-Saharan African ethics (available in the literature) might sit in direct opposition to the fundamental motivation of corporate adoption of artificial intelligence; these values are Afro-communitarianism grounded on relationality, and human dignity grounded on a normative conception of personhood. This paper offers a unique perspective to AI ethics from the African place, as there is little to no material in the literature that discusses the implications of AI on African ethical values. The paper is divided into two broad sections that are focused on (i) describing the values at risk from AI and (ii) showing how current use of artificial intelligence undermines these said values. In conclusion, I suggest how to prioritize these values in working toward the establishment of an African AI ethics framework.

Artificial Intelligence, Digital Colonialism and the Implications for Africa's Future Development

Aishat **Salami**
Technology Consulting and Research, Veeta Advisory Hub, Lagos, Nigeria

Sub. No : DAP-2023-0174 – Sp3
Full paper is accepted for publication in the Data & Policy journal.

Abstract

In the mid to late 19th century, much of Africa was under colonial rule, with the colonisers exercising power over the labour and territory of Africa. However, as much as Africa has predominantly gained independence from traditional colonial rule, another form of colonial rule still dominates the African landscape. This similitude of these different forms of colonialism is found in the power dominance exhibited by Western technological corporations, just like the traditional colonialists. In this digital age, digital colonialism manifests in Africa through the control and ownership of critical digital infrastructure by foreign entities, leading to unequal data flow and asymmetrical power dynamics. This usually occurs under the guise of foreign corporations providing technological assistance to the continent.

By drawing references from the African continent, this paper examines the manifestations of digital colonialism and the factors that aid its occurrence on the continent. It further explores the manifestations of digital colonialism in technologies such as Artificial Intelligence while analysing the occurrence of data exploitation on the continent and the need for African ownership in cultivating the digital future of the African continent. The paper also recognises the benefits linked to the use of Artificial Intelligence and makes a cautious approach towards the deployment of AI tools in Africa. It then concludes by recommending the implementation of laws, regulations, and policies that guarantee the inclusiveness, transparency, and ethical values of new technologies, with strategies towards achieving a decolonised digital future on the African continent.

Should we communicate with the dead to assuage our grief? An Ubuntu perspective on using griefbots

Connor **Wright**[1,2]

1. *LCFI, University of Cambridge, Cambridge, UK*
2. *Montreal AI Ethics Institute, Montreal, Canada*

Sub. No: DAP-2023-0141– Sp3
Full paper is accepted for publication in the Data & Policy journal.

Abstract

During the 20th century, dealing with grief through an on-going involvement with the deceased (such as speaking to their grave) was seen as pathological by Western authors such as Sigmund Freud. Nowadays, we are presented with the opportunity to continue interacting with digital representations of the deceased. As a result, the paper adopts an Ubuntu perspective, i.e., a sub-Saharan African philosophy focussed on community and relationship to provide a toolkit for using this emerging technology. I will argue that the Ubuntu framework I propose contributes to the use of griefbots in two ways. The first, is that it shows that it is morally permissible to use griefbots to assuage our grief. The second, is that it delineates how we can ethically use the technology. To do so, I split my analysis into four sections. In the first section, I show that meaningful relationships can occur between the bereaved and griefbots. This will be done by exploring the Western theory of continuing bonds proposed by Dennis Klass, Phyllis Silverman and Steven Nickman. In my second, I flesh out my Ubuntu framework according to Thaddeus Metz's accounts on Ubuntu as a modal-relational theory. In my third section, I apply my Ubuntu framework to the case of Roman Mazurenko. Furthermore, I consider some counterarguments to the Ubuntu framework in terms of how it could lead the user to encounter issues surrounding privacy, commercialisation and people replacement. Finally, I conclude that, despite these limitations, the Ubuntu framework positively contributes to determining whether we should communicate with the dead through griefbots to assuage our grief.

The Ethics at the Intersection of Artificial Intelligence and Transhumanism: A Personhood-Based Approach

Amara Esther **Chimakonam**
Centre for Phenomenology in South Africa, University of Fort Hare, Alice, South Africa

Sub. No: DAP-2023-0132 – Sp3
Full paper is accepted for publication in the Data & Policy journal.

Abstract

In this paper, I will consider the moral issues that might arise from the possibility of creating more complex and sophisticated autonomous intelligent machines or simply Artificial Intelligence (AI) that would have the human capacity for moral reasoning, judgment, and decision-making, and [the possibility] of humans enhancing their moral capacities beyond what is considered normal for humanity. These two possibilities raise an urgency for ethical principles that could be used to analyze the moral consequences of the intersection of AI and transhumanism. In this paper, I deploy personhood-based relational ethics grounded on Afro-communitarianism as an African ethical framework to evaluate some of the moral problems at the intersection of AI and transhumanism. In doing so, I will propose some Afro-ethical principles for research and policy development in AI and transhumanism

Responsible artificial intelligence in Africa: Towards policy learning

Paul **Plantinga**[1], Kristophina **Shilongo**[2], Oarabile **Mudongo**[3], Angelique **Umubyeyi**[4], Michael **Gastrow**[1] and Gabriella **Razzano**[5]

1. *Human Sciences Research Council, South Africa*
2. *Mozilla Foundation, Namibia*
3. *Consumers International, Botswana*
4. *Independent, South Africa*
5. *OpenUp, South Africa*

Sub. No: DAP-2023-0133 – Sp3 (Remote Presentation)
Full paper is accepted for publication in the Data & Policy journal.

Abstract

Several African countries are developing artificial intelligence (AI) strategies and ethics frameworks with the goal of accelerating responsible AI development and adoption. However, many of these governance actions are emerging without consideration for their suitability to local contexts, including whether the proposed policies are feasible to implement and what their impact may be on regulatory outcomes. In response, we suggest that there is a need for more explicit policy learning, by looking at existing governance capabilities and experiences related to algorithms, automation, data and digital technology in other countries and in adjacent sectors. From such learning it will be possible to identify where existing capabilities may be adapted or strengthened to address current AI-related opportunities and risks. This paper explores the potential for learning by analysing existing policy and legislation in twelve African countries across three main areas: strategy and multi-stakeholder engagement, human dignity and autonomy, and sector-specific governance. The findings point to a variety of existing capabilities that could be relevant to responsible AI; from existing model management procedures used in banking and air quality assessment, to efforts aimed at enhancing public sector skills and transparency around publicprivate partnerships, and the way in which existing electronic transactions legislation addresses accountability and human oversight. All of these point to the benefit of wider engagement on how existing governance mechanisms are working, and on where AI-specific adjustments or new instruments may be needed.

Trust Norms for Generative AI Data Gathering in the African Context

Abiola Joseph **Azeez**[1] and Tosin **Adeate**[2]

1. *Philosophy Department & Canadian Robotics and Artificial Intelligence Ethical Design Laboratory, University of Ottawa, Canada*
2. *Department of Philosophy, Olabisi Onabanjo University, Nigeria*

Sub. No: DAP-2023-0169 – Sp3
Full paper is accepted for publication in the Data & Policy journal.

Abstract
Can trust norms within the African moral system support data gathering for Generative AI (GAI) development in African society? Recent developments in the field of large language models, such as GAI, including models like ChatGPT and Midjourney have identified a common issue with these GAI models known as "AI hallucination," which involves the presentation of misinformation as facts along with its potential downside of facilitating public distrust in AI performance. In the African context, this paper frames unsupportive datagathering norms as a contributory factor to AI hallucination and investigates the following claims. Firstly, this paper explores the claim that knowledge in the African context exists in both esoteric and exoteric forms, incorporating such diverse knowledge as data could imply that a GAI tailored for Africa may have unlimited accessibility across all contexts. Secondly, this paper acknowledges the formidable challenge of amassing a substantial volume of data, which encompasses esoteric information, requisite for the development of a GAI model, positing that the establishment of a foundational framework for data collection, rooted in trust norms that is culturally resonant, has the potential to engender trust dynamics between data providers and collectors.

Lastly, this paper recommends that trust norms in the African context require recalibration to align with contemporary social progress, while preserving their core values, to accommodate innovative data-gathering methodologies for a GAI tailored to the African setting. This paper contributes to how trust culture within the African context, particularly in the domain of GAI for African society, propels the development of Afro-AI technologies.

Case Studies of AI Policy Development in Africa

Kadijatou **Diallo**[1], Jonathan **Smith**[2], Chinasa T. **Okolo**[3], Dorcas **Nyamwaya**[4], Jonas **Kgomo**[4] and Richard **Ngamita**[4]

1. *Harvard Kennedy School, Harvard University, Boston, Massachusetts, United States*.
2. *Meta, Menlo Park, California, United States*
3. *Center for Technology Innovation, The Brookings Institution, Washington D. C., United States*
4. *Equiano Institute, Nairobi, Kenya*

Sub. No: DAP-2023-0177 – Sp3
Full paper is accepted for publication in the Data & Policy journal.

Abstract

Artificial Intelligence (AI) requires new ways of evaluating national technology use and strategy for African nations. We conduct a survey of existing 'readiness' assessments both for general digital adoption and for AI policy in particular. We conclude that existing global readiness assessments do not fully capture African states' progress in AI readiness and lay the groundwork for how assessments can be better used for the African context. We consider the extent to which these indicators map to the African context and what these indicators miss in capturing African states' on-the-ground work in meeting AI capability. Through case studies of four African
nations of diverse geographic and economic dimensions, we identify nuances missed by global assessments and offer high-level policy considerations for how states can best improve their AI readiness standards and prepare their societies to capture the benefits of AI.

Social Justice Considerations in Developing and Deploying AI in Africa

Getachew Hailemariam **Mengesha**[1], Elefelious Getachew **Belay**[2] and Rachel **Adams**[3]

1. *School of Information Science, Addis Ababa University, Addis Ababa, Ethiopia*
2. *School of Information Technology and Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia*
3. Global Center on AI Governance

Sub. No: DAP-2023-0185 – Sp3 (Remote Presentation)
Full paper is accepted for publication in the Data & Policy journal.

Abstract

In the literature, there are polarized views regarding the capabilities of technology to embed societal values. One aisle of the debate contends that technical artifacts are value-neutral since values are not peculiar to inanimate objects. Scholars on the other side of the aisle argue that technologies tend to be value-laden. With the call to embed ethical values in technology, this paper explores how AI and other adjacent technologies are designed and developed to foster social justice. Drawing insights from prior studies, this paper identifies seven African moral values considered central to actualizing social justice; of these, two stand out — respect for diversity and ethnic neutrality. By introducing use case analysis along with the Discovery, Translation, and Verification (DTV) framework and validating via Focus Group Discussion, this study revealed novel findings: firstly, ethical value analysis is best carried out alongside software system analysis. Secondly, to embed ethics in technology, interdisciplinary expertise is required. Thirdly, the DTV approach combined with the software engineering methodology provides a promising way to embed moral values in technology. Against this backdrop, the two highlighted ethical values – respect for diversity and ethnic neutrality – help ground the pursuit of social justice.

AI for Women's Financial Inclusion – Analysis of Product Design and Policy Approaches in Nigeria

Adekemi **Omotubora**[1]

1. *Department of Commercial and Industrial Law, University of Lagos, Akoka, Nigeria*

Sub. No: DAP-2023-0144
Full paper is accepted for publication in the Data & Policy journal.

Abstract

Nigeria has a significant gender financial inclusion gap with women disproportionately represented among the financially excluded. Artificial intelligence (AI) powered financial technologies (fintech) present distinctive advantages for enhancing women's inclusion. This includes efficiency gains, reduced transaction costs, and personalized services tailored to women's needs. Nonetheless, AI harbours a paradox. While it promises to address financial inclusion, it can also inadvertently perpetuate and amplify gender bias. The critical question is thus, how can AI effectively address the challenges of women's financial exclusion in Nigeria? Using publicly available data, this research undertakes a qualitative analysis of AI-powered Fintech services in Nigeria. Its objective is to understand how innovations in financial services correspond to the needs of potential users like unbanked or underserved women. The research finds that introducing innovative financial services and technology is insufficient to ensure inclusion. Financial inclusion requires the availability, accessibility, affordability, appropriateness, sustainability, and alignment of services with the needs of potential users, and policy driven strategies that aid inclusion.

Knowledge, Attitudes and Readiness Towards Artificial Intelligence in Government Services; A developing Country Perspective

Eric Afful-**Dadzie**[1] and Samuel **Lartey** [1]

1. University of Ghana Business School

Sub. No: 2122 – Sp3

Abstract

Artificial Intelligence (AI) is a pervasive technology that is increasingly employed in many sectors across the globe. In developing countries like Ghana, its application is being championed by financial and telecommunication organizations whereas the public sector (government services) remains a gaping distance behind in its adoption and utilization. Whiles this presents the current reality, there is evidence of AI being a central player of government services in the coming future. As Ghana prepares towards this impending future, it is important to gauge the perceptions of citizens towards the use of AI in government services. The study therefore employs configurational techniques, particularly fuzzy set qualitative comparative analysis (FSQCA) to unearth the diversity in user types that will influence the successful use of these government-initiated AI services. In essence, this study seeks to provide an answer to the question; how can government-initiated AI services be successfully accepted by the citizenry? Leveraging a random sample of 385 tertiary educated students from across the country, the study utilizes a survey to solicit respondents' knowledge, attitudes, readiness, and use intentions towards AI-enabled government services. The FSQCA results suggest four unique profiles of citizens that are likely to embrace the use of these services. The first profile of users are those that are knowledgeable in the use of AI and are also innovative in AI. These same users however have a non-positive attitude towards AI, do not show discomfort in such technologies and also have no insecurities with using the technology. The second profile of users are knowledgeable, innovative and optimistic about AI technologies in government services and also have a positive attitude towards AI in government services. This user profile also do not show discomfort in AI use. The third profile of users are those that have a positive attitude towards AI in government services, are optimistic and innovative, but are however less knowledgeable in AI technologies and show discomfort with such services. The final user profile that was determined are those that have knowledge of AI in government services, are optimistic and have a positive attitude towards the technology. However, these same users show a level of discomfort with the use of the technology in government services and also have some level of insecurity with its use. An analysis of the core factors that contribute to citizenry acceptance of AI in government services is their knowledge of technology, their optimism towards its use and their innovative abilities.

From the findings, the study recommends that, government agencies seeking to implement AI to complement their services must be intentional in building the citizens knowledge about these technologies and creating optimism around its use. They must also pay attention to how users can innovatively use these technologies. Additionally, it is important to make necessary efforts to assuage citizens discomfort and insecurities with the use of AI in government services. The findings are essential in driving the efficient implementation of AI-enabled services amongst a broad group of educated members of the citizenry in developing country contexts.

Towards a Fair and Equitable Data Ecosystem for Low Resource languages.

Jonas **Kgomo**[1], Dorcas **Nyamwaya**[1], Abigail **Oppong**[1], Chinasa T. **Okolo**[1] and Susan **Otieno**[1]

1. *Equiano Institute, Nairobi, Kenya*

Sub. No: 9826 – Sp3

Abstract

There are several potential strategies that could be adopted to achieve a fair and equitable data ecosystem for Low Resource Languages, including open sourcing the data, implementing licensing agreements, and establishing closed-source API-based access with added security measures. Each approach has its limitations, with open sourcing being vulnerable to exploitation by bad actors, licensing potentially leading to loss of control through poor monitoring, and closed-source access raising concerns around distribution of benefits, data privacy and consent.
Additionally, the economic implications of acquiring LRL data pose challenges in determining appropriate value and usage, given the potential externalities and community concerns. In order to devise an effective and ethical solution to this problem, it is crucial to explore innovative data governance strategies that balance the interests of the LRL communities and the broader goal of improving large language models for these languages.

This paper explores the role of data governance in Artificial Intelligence, highlighting the challenges and opportunities of data governance in this context, and proposes a framework for developing data governance policies that are tailored to the needs of the global south.

Community Based AI Governance

Jonas **Kgomo**
Equiano Institute, Nairobi, Kenya

Sub. No: 4596

Abstract

This paper examines how crowdsourced data initiatives can enhance oversight[1], evaluation, and red teaming[2] around governance issues in Africa. We analyze cases where mobile technology and digital civil society groups empower citizens to document conditions around them related to topics like electoral fraud, corruption, public service delivery, outbreaks of violence, and environmental damage. These crowd-enabled reporting systems act as a form of participatory LLMs, unveiling critical governance gaps around security, justice, funding allocations, and administrative performance. We spotlight citizen reporting platforms in Kenya, Nigeria, and South Africa focused specifically on integrity issues in legal and judicial processes. The rise of these crowdsourced monitoring and transparency efforts has pressed African governments toward more accountability, showcasing civic data's value in applied policy learning and anti-corruption reforms.

In addition to these grassroots efforts, we started The Africa Oversight TAO186 Coalition, a pioneering AI Safety Coalition set to launch in 2024. TAO is dedicated to advancing the fairness and integrity of AI technologies in the Global South, with an inaugural focus on African nations. Comprising experts from organizations, government bodies, academia, and industry leaders, TAO aims to inform the steerable development and deployment of AI technologies, including Large Language Models LLMs).

TAO's evaluation methods include Red Teaming Exercises to conduct controlled adversarial attacks, Model API Evaluation for rigorous testing of AI model performance, and Human-Led Assessments to judge model fairness and mitigate bias. Furthermore, TAO will lead the development of a comprehensive framework outlining best practices for the ethical development and deployment of AI models within the African context. This framework will be regularly updated and disseminated to African governments and businesses, serving as a guiding resource for responsible AI utilization.

We analyze the potential for crowdsourced data initiatives to enhance oversight, red teaming, and evaluations around governance in Africa. We examine cases where digital civil society groups and grassroots reporting networks create participatory monitoring systems to document issues in public integrity, service delivery, electoral processes, and administration. These crowd-enabled platforms act as a form of distributed oversight and rapid feedback around critical governance gaps impacting security, legal rights, funding equity, and state performance.

However, open citizen reporting also enables “red teaming” from both civil society and institutional reformers to pressure-test response protocols, verify data reliability, and counter potential manipulation or sampling bias. Constructively integrating crowdsourced oversight with formal policy frameworks can optimize applied learning while mitigating risks. Our analysis weighs data quality control tradeoffs and proposes structural options for public agencies to leverage crowd insights through embedded evaluative processes like automated red team audits, participatory policy formulation, and collaborative anti-fraud reforms while preserving institutional independence. We conclude by spotlighting innovative hybrid civil society/governmental oversight models and evaluating their reform impact.

Bibliography

1187 The Africa OversightTAO187 by Jonas Kgomo
2187 GitHub - equiano-institute/haystack: A suite of red teaming and evaluation frameworks for language models by Jonas Kgomo

Social Media and Government

A political economy of information disorder in South and Southeast Asia

Nicola **Nixon**
The Asia Foundation

Sub. No. 3059 – Sp4

Abstract

Digital technology has revolutionized the pace, cost, and sources of information creation and distribution. As Asia transitions to information societies, the region faces risks to political processes and social relations from the level of information distortion and manipulation it is experiencing. The digital public sphere has moved beyond the phase of instances of 'fake news'. Mis and disinformation, spread at high speed and across vast distances by those who profit from it, interact with political systems to exacerbate divisions, increase social polarization and cause violence.

In this paper, we draw on documentary sources and interviews with civil society organizations, to examine the politics of this phenomenon in Indonesia, the Philippines and Sri Lanka and the efforts of those at the frontline who are trying to respond.

While propaganda and hate speech are not new phenomena in these countries - as they are not elsewhere – their experience of communication production and consumption shares some similarities to that South and Southeast Asia more broadly. Our analysis looks at digital public spheres in contexts in which colonial and authoritarian histories were underpinned by highly stratified media landscapes. That lends itself to higher levels of trust in social media, seen as alternative.

In terms of production trends, we look at the array of actors involved, during elections and more broadly. Information disorder during elections has evolved in all three countries, from troll farm operations run by public relations firms, macro and micro influencers, 'buzzers' and other gig workers on short term contracts, and the politicians and political parties who tacitly or explicitly support them. In the most recent elections in the Philippines and Indonesia, this has evolved into 'cyber manipulation' which no longer necessarily involves blatantly fake news but a more complex mix of truth and falsehoods that is more difficult to challenge.

Underpinning the disorder in election information is the ongoing proliferation of disinformation that targets ethnic and religious minorities and women. Coupled with some form of incitement to harassment of ethnic and religious minorities, mis and disinformation stokes tensions and causes conflict.

In terms of consumption, it is important to understand internet usage patterns. In all three countries, most people use their mobile phones – rather than computers or tablets – to access the internet. This means, for many people Facebook is the internet. Coupled with the fact that most traditional media sources are paywalled, and that so much of the web is in

languages other than those in which they are fluent, few internet users in the region have access to the internet as a vast source of information in the way users do in the Global North.

Given this context, several groups emerge as particularly vulnerable to disinformation. This includes older people and children who tend to have lower digital literacy and a greater susceptibility to conspiracy theories and blatantly false narratives. It also includes what in Indonesia has been called 'the scooter class' and in the Philippines is considered the 'precarious non-poor': those whom, during the country's transition to middle income status, can easily backslide into poverty if they experience an economic or health shock. Where they also live in communities where crime rates are high, they are more receptive to messages around safety and security framed in nationalist populist terms.

The impact of information disorder in Indonesia, the Philippines and Sri Lanka is multifaceted. It stretches from the violence experienced by individuals and groups, through negative impacts on the integrity of election processes, to more broadly weakening social cohesion in countries with weak governance and many, many development challenges. At the individual and community level, recent research in Sri Lanka, found that since the end of the conflict in 2009, hate speech disseminated on social media has contributed to violence against several minority communities. After the Easter bombings in 2019, for instance, online hate speech contributed to an increase in violence towards Muslim minorities. Similarly, research in Indonesia has found a disproportionate impact on minority communities including psychological stress, economic damage, and sexual and physical violence. Those impacted include women, sexual minorities, ethnic minorities and persons with disabilities. In the Philippines, attacks on human rights advocates – a trend that started during the Duterte administration – are an issue of particular concern.

Currently, the most prominent response mechanism in the region is fact-checking. Dozens of organizations are involved in independent, third party fact-checking of social media posts throughout the region. Many are registered with platforms such as Facebook with whom they work to remove misleading and false posts.

Our research suggests that fact-checking is useful. Posts that are fact-checked tend to be shared less by users. Fact-checking helps to raise awareness about the impact of mis and disinformation. With the increased use of AI to spread disinformation, fact-checking is a necessary but limited response in contexts in which there are few alternatives.

Although governments that perpetuate disinformation may be unlikely to play a constructive role in regulating or controlling it, that doesn't mean policy responses are dismissed altogether. Electoral laws and systems, for instance, are being updated to accommodate the realities of social media use. Many also recognize how important education systems are in long-term responses to the political and social impacts of information disorder. Where there are government champions willing to go against the grain, some of these efforts may bear fruit.

Our research shows that civil society organizations in Indonesia, the Philippines and Sri Lanka have developed sophisticated responses to the disinformation storm. Civil society recognize that this is more than an 'internet' issue, but one that jeopardizes social cohesion and democratic governance more broadly. They are, however, sorely under-resourced and operating in contexts in which the broader political economy is not conducive to change,

and where politicians and those in power are not only benefitting from information disorder but are colluding with private companies to produce it in the first place.

Social Media in Elections: A Glimpse of Mis/Disinformation from Developing Countries

Charmaine **Distor**[1]**,** Danilo **Đikanović**[1], Soumaya **Ben Dhau**[1]

1. *United Nations University*

Sub. No*:* 9103 – Sp4

Abstract

Technology, particularly social media, plays an essential role in shaping societal perceptions and political discourse in the contemporary political landscape. Scholars like Hendricks and Schill (2017) assert that modern political campaigns are integrally interlinked with social media use, highlighting their critical role in contemporary political communication strategies. Social media's significance lies not only in its role as a platform for information dissemination but also in its capacity to facilitate engagement and community-building. For instance, Obama's 2008 campaign leveraged social media platforms to disseminate information, gather data and strategically foster virtual communities (Cogburn & Espinoza-Vasquez, 2011). The accessibility, interactivity, and immediacy of social media have transformed the landscape of political communication, with contextual factors such as actors, timing, and political systems further amplifying its influence (Zhang, 2016). While existing literature extensively covers the potential benefits of social media for political campaigning, including its cost-effectiveness, lack of traditional media gatekeeping, and impact on electoral outcomes (Strandberg, 2013; Hendricks & Schill, 2017; Gibson, 2015; Smyth & Best, 2013; Brito et al., 2019; Bright et al., 2020), research on its adverse effects, particularly within the democratic frameworks of developing economies, remains rare. Hence, this study aims to contribute to this literature on the intersection of social media and governmental affairs, focusing specifically on the prevalence of mis/disinformation during electoral periods in developing countries. The main objective is to analyse the strategies used by the key stakeholders within this ecosystem and assess their effects on electoral processes.

One adverse effect is the misuse of personal data from social media platforms, exemplified by the Cambridge Analytica scandal, which aimed to manipulate voters' behaviour in significant electoral events such as the US presidential election and the Brexit referendum (Confessore, 2018; Hinds, Williams, & Joinson, 2020). Additionally, social media's propensity to blur the lines between fact and opinion, news and entertainment, and information producers and consumers has been evident in recent elections globally, notably during the 2016 US Presidential election (Delli Carpini, 2016; Hendricks & Schill, 2017). These issues underscore the ominous aspects of social media's influence on elections, including concerns about data privacy and information manipulation, particularly

within developing countries where social media plays an increasingly pivotal role in shaping political landscapes.

This paper investigates the phenomenon of mis/disinformation during elections in three developing countries—namely, the Philippines, Montenegro, and Tunisia—through a qualitative approach encompassing a literature review, case studies, and comparative analysis. These countries were selected for their diverse cultural contexts, governance models, and geographic coverage to yield widely applicable findings. The analysis identifies the influential actors within the mis/disinformation ecosystem, explores their political effects, and examines initiatives to prevent them. Based on the findings, policy and research recommendations are provided to mitigate the adverse impacts of mis/disinformation perpetrated by influential actors during elections.

This study demonstrated the importance of investigating the influential actors in the mis/disinformation ecosystem. The results showed that understanding who the actors are and how they perpetuate mis/disinformation is essential for preserving the integrity of electoral processes. The results validate past studies that have already argued the critical role that social media users play in magnifying political mis/disinformation (Dupuis & Williams, 2019) and state-sponsored trolls in manipulating public opinion through coordinated campaigns (Zannettou et al., 2018). It also emphasizes Zhang's thesis (2016) that social media shaping public opinion is significantly influenced by various factors, encompassing key actors, timing, and the prevailing political context.

The study concluded the prominence of various actors in disseminating mis/disinformation, accentuated by the urgency and importance of the electoral period. Notably, the democratic fabric characterizing governance in the Philippines, Montenegro, and Tunisia has been fertile ground for the proliferation of mis/disinformation. Within this context, political aspirants in these nations leverage social media platforms for electoral campaigns. However, the absence or obsolescence of electoral campaign regulations has led to the negligent or malevolent use of social media channels. Political candidates and public officials have disseminated false information, even in official statements and venues, which both trustworthy and questionable media sources amplified, spreading further on social media. Noteworthy among the outlets of mis/disinformation are not only political candidate supporter pages but also voluntary citizens, paid influencers, troll farms, and questionable media entities.

Across the electoral landscapes of the three countries, a pattern emerges regarding the typology of mis/disinformation, manifesting primarily as attacks on electoral rivals. This strategy, anchored in historical political campaigns, seamlessly extends into the digital sphere. Moreover, mis/disinformation assumes diverse forms, including manipulating historical narratives and disseminating questionable research findings. The repercussions of mis/disinformation during electoral cycles are profound and manifold. Apart from tarnishing the reputations of targeted candidates, it has also facilitated the electoral successes of certain politicians. Furthermore, it has increased distrust in governmental institutions and credible media outlets, encouraging influential actors to fortify the mis/disinformation ecosystem. In extreme cases, the dissemination of mis/disinformation has caused threats against journalists, influenced voter behaviour negatively, and even incited violence.

Despite the persistent spread of mis/disinformation, countermeasures are gradually gaining traction, particularly during electoral periods. Foremost is the proliferation of fact-checking initiatives, predominantly led by reputable media entities. Collaborative efforts among various stakeholders—comprising governments, NGOs, media outlets, and academic institutions—have also fostered dialogues and initiatives to enhance media literacy and combat mis/disinformation.

Nevertheless, concerted actions are imperative to address this issue effectively. Strategic policy interventions, including comprehensive disclosure requirements for social media expenditures during political campaigns and robust research, are required. Collaboration among stakeholders, encompassing governments, tech giants, startups, NGOs, academia, and media outlets, holds promise for innovative solutions. Moreover, strengthening voter education and media literacy initiatives, alongside exploring emerging technologies like AI and open data, are crucial steps toward mitigating the adverse effects of mis/disinformation. Regulatory frameworks must be agile and responsive to the evolving digital landscape, highlighting the importance of constantly adapting to safeguard democratic processes and public discourse.

References

Bright, J., Hale, S., Ganesh, B., Bulovsky, A., Margetts, H., & Howard, P. (2020). Does Campaigning on Social
Media Make a Difference? Evidence from candidate use of Twitter during the 2015 and 2017 UK Elections.
Communication Research, *47*(7), 988-1009. https://doi.org/10.1177/0093650219872394.

Brito, K., Meira, S., Paula, N., & Fernandes, M. (2019). Social media and presidential campaigns–preliminary
results of the 2018 Brazilian presidential election. In *dg.o 2019: 20th Annual International Conference on*
Digital Government Research, June 18–20, 2019, Dubai, United Arab Emirates (pp. 332-341). New York,
USA: ACM. https://doi.org/10.1145/3325112.3325252.

Cogburn, D. L., & Espinoza-Vasquez, F. K. (2011). From networked nominee to networked nation: Examining
the impact of Web 2.0 and social media on political participation and civic engagement in the 2008 Obama
campaign. *Journal of Political Marketing*, *10*(1-2), 189-213.
https://doi.org/10.1080/15377857.2011.540224.

Confessore, N. (2018). Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. Available at
https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

Delli Carpini, M. X. (2016). The new normal? Campaigns & elections in the contemporary media environment.
US Election Analysis 2016: Media, Voters and the Campaign. Available at
http://www.electionanalysis2016.us/us-election-analysis-2016/section-1-media/the-new-normal-campaigns-

elections-in-the-contemporary-media-environment/.
Dupuis, M., & Williams, A. (2019). The Spread of Disinformation on the Web: An Examination of Memes on
Social Networking. *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted
Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and
Smart City Innovation,* 1412-1418. https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-
SCI.2019.00256
Gibson, R. K. (2015). Party change, social media and the rise of 'citizen-initiated'campaigning. *Party Politics*,
21(2), 183-197. https://doi.org/10.1177/1354068812472575.
Hendricks, J. A., & Schill, D. (2017). The social media election of 2016. In R. E. Denton Jr. (ed.), *The 2016 US
Presidential Campaign. Political Campaigning and Communication*, (pp.121-150). Palgrave Macmillan, Cham.
https://doi.org/10.1007/978-3-319-52599-0_5.
Hinds, J., Williams, E. J., & Joinson, A. N. (2020). "It wouldn't happen to me": Privacy concerns and
perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*,
143, 102498. https://doi.org/10.1016/j.ijhcs.2020.102498.
Smyth, T. N., & Best, M. L. (2013). Tweet to trust: Social media and elections in West Africa. In *ICTD '13:
Proceedings of the Sixth International Conference on Information and Communication Technologies and
Development, December 7–10, 2013, Cape Town, South Africa* (pp. 133-141). New York, USA: ACM.
https://doi.org/10.1145/2516604.2516617.
Strandberg, K. (2013). A social media revolution or just a case of history repeating itself? The use of social
media in the 2011 Finnish parliamentary elections. *New Media & Society*, *15*(8), 1329-1347.
https://doi.org/10.1177/1461444812470612.Zannettou, S., Caulfield, T., Setzer, W., Sirivianos, M., Stringhini, G., & Blackburn, J. (2018). Who Let The
Trolls Out?: Towards Understanding State-Sponsored Trolls. *Proceedings of the 10th ACM Conference on Web
Science, 19, 353–362*. https://doi.org/10.1145/3292522.3326016.
Zhang, W. (2016). Social media and elections in Singapore: comparing 2011 and 2015. *Chinese Journal of
Communication*, *9*(4), 367-384. https://doi.org/10.1080/17544750.2016.1231129.

Generative AI for Sound Decision-making

Enhancing Health Policy-Making Through ChatGPT: Opportunities and Threats

Shahabeddin **Abhari**[1], Plinio **Morita**[1] and Jasleen **Kaur** [1]

1. *School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada*

Sub No: 4258 – Sp6

Abstract

Introduction

Health policy-making grapples with complexities amid the COVID-19 pandemic, an aging population, and rising healthcare costs. The integration of artificial intelligence (AI), particularly ChatGPT, in 2022, introduces novel avenues for innovation in health policy-making. The aim of this perspective study is to delve into the applications of AI bots, with a particular focus on ChatGPT, and assess their potential contributions to the field of health policy-making.

Method

In order to comprehensively explore the topic, we conducted a thorough literature review, utilizing respected databases such as Pubmed, Scopus, Embase, and Google Scholar. Our search involved specific keywords, including "ChatGPT" or "Natural Language Processing" combined with terms like "Health Policies," "Health Policy," "Healthcare Policy," "National Health Policy," or "Policy Making" to refine the results. The search was limited to English-language papers published before April 10, 2023, to ensure precision. After identifying relevant articles, a careful evaluation led to the selection of specific papers for in-depth analysis. Subsequently, we employed qualitative content analysis and a research panel to extract and categorize key insights related to ChatGPT's potential applications, opportunities, and challenges in the field of health policy-making.

Results

Applications of ChatGPT in Health Policy-Making

ChatGPT, a language model by OpenAI, streamlines communication, facilitates data analysis, and provides insights. Applications include policy analysis and development, policy communication, policy evaluation, decision-making support, and bias identification.

Policy Analysis and Development: ChatGPT can revolutionize policy analysis by automating evidence-based recommendations. During a pandemic, it could analyze global healthcare policies and research data to recommend optimal vaccination strategies.

Policy Communication: ChatGPT enhances transparency by generating plain-language summaries of complex policy documents, fostering understanding, especially for new policies like telemedicine.
Policy Evaluation: ChatGPT's data analysis capabilities offer valuable insights into policy effectiveness, such as evaluating the success of a nationwide smoking cessation campaign.
Decision-Making: ChatGPT supports real-time decision-making by providing responses and identifying trends, enabling informed decisions in critical areas.
Bias Identification and Mitigation: ChatGPT contributes to fairness by identifying and mitigating bias, flagging inconsistencies across demographic groups.
Opportunities and Threats
The integration of ChatGPT in health policy-making presents opportunities and threats.

Opportunities

Improved Efficiency: AI bots enhance efficiency by automating tasks, reducing time and resources for manual data analysis.
Enhanced Decision-Making: AI bots support decision-making by providing data-driven insights, informing critical areas of healthcare decision-making.
Increased Access: AI bots engage stakeholders, providing personalized health recommendations and advice based on health data.
Cost Savings: By automating tasks like data analysis, AI bots reduce costs associated with health policy-making.

Threats

Data Privacy Concerns: AI bots' reliance on extensive healthcare data raises concerns about data privacy breaches.
Algorithmic Bias: Training AI bots on biased data may result in biased recommendations, perpetuating disparities in healthcare.
Potential for AI to Replace Human Decision-Making: Concerns exist about reducing human decision-making roles and a lack of transparency in AI-generated recommendations.
Ethical Concerns: The use of AI bots raises ethical concerns, necessitating transparent regulatory frameworks for fairness and accountability.

Conclusion:

In conclusion, the integration of ChatGPT and other AI bots in health policy-making offers opportunities such as improved efficiency and reduced workload for healthcare professionals. However, potential threats include data privacy concerns, algorithmic bias, and the risk of AI replacing human decision-making. Future investigations should assess the practical benefits. Policymakers need to carefully consider these challenges, prioritizing ethical and responsible AI use. This involves ensuring data security, minimizing algorithmic bias through diverse training datasets, and promoting collaboration with healthcare professionals. While ChatGPT and AI bots present an opportunity to enhance healthcare outcomes, a cautious approach is crucial to maximize benefits and mitigate risks, ensuring ethical and impactful integration into health policy-making processes.

Generative AI and electric vehicle service operations in urban and remote areas

Omar **Asensio**[1] and Yifan **Liu**[2]

1. *Harvard Business School*
2. *Georgia Institute of Technology*

Sub. No: 9951 – Sp6

Abstract

For the first time in nearly three decades, the transportation sector is now the largest source of U.S. greenhouse gas emissions. To accelerate climate action, governments are promoting zero emission vehicles (ZEV) policies to accelerate the electrification of cars and trucks, as well as increase equity in access to public charging facilities. However, given the decentralized models of charging station growth, individual station operators set prices and access policies, which have created data interoperability challenges for large-scale analysis of service operations. In this talk, I will describe the use of generative AI and expert specialization to overcome fundamental evaluation challenges with distributed and unstructured digital consumer data, particularly in the context of electric vehicles. By guiding context learning with chain-of-thought prompting, we significantly reduce research evaluation costs with GPT-4, compared with conventional methods of analysis. Using this approach, we evaluate the state of the U.S. electric vehicle charging infrastructure from 2011-2022. The analysis covers 31,527 chargers nationwide, with special emphasis on reliability and distributive-equity issues that impact climate-disadvantaged communities. We uncover evidence that failures in service operations are dominant challenges to the delivery of public charging services. Survival analysis also indicates significantly lower survival rates for charging stations located in climate-disadvantaged communities, particularly for those that are not part of a network with contractual maintenance subscription services. Non-networked stations also face a higher frequency of station losses by consumers. Evidence shows persistent reliability and service provision gaps affecting 22.7 million individuals, particularly in rural communities and urban clusters not expected to be in targeted federal investment zones.

Does Generative AI Revolutionize Higher Education? Perspectives, Policies, and Curriculum Reforms in Top Asian Universities

Wilson **Wong**[1], Angela **Aristidou**[2], Konstantin **Scheuermann**[2] and Tony **Wong**[1]

1. *The Chinese University of Hong Kong*
2. *UCL*

Sub. No: 8502 – Sp6

Abstract

This paper assesses the impact and transformation of Generative Artificial Intelligence (AI) particularly ChatGPT on higher education with a focus on Asian universities. It would address the main research questions of both theory development and empirical analysis. In the former, building on a comprehensive literature review, it would then develop a theoretical framework to identify the challenges for higher education in the AI era in terms of learning and curriculum and the strategic leadership, organizations, and institutions required in the transformation. In the latter, it would examine through empirical analysis of perspectives, policies and curriculum reforms in top universities in Asian studies. Those cases would represent a benchmark for comparison of regions across the East and the West including North America, Asia and Europe. Through the empirical analysis of perspectives, policies, and curriculum reforms of the top universities in Asia, it will examine the existence of any gap between theory and practice.

Generative Artificial Intelligence (GAI), including transformative technologies such as ChatGPT, is rapidly changing the contours of various sectors of human life. One domain standing at the precipice of this monumental shift is higher education. As we navigate the threshold of an era where AI technologies possess the power to redefine our traditional learning and teaching methodologies, some critical questions arise: What capacities should we offer to students of higher education and what curriculum reforms are needed accordingly? How prepared are our educational institutions to embrace this shift? This paper embarks on an exploration of this pressing issue, with a concentrated focus on the role and readiness of top Asian universities in the face of the rising tide of GAI.

The advent of GAI presents a dual challenge for universities worldwide. The first challenge is understanding and adaptation: educational institutions must find ways to comprehend the meanings and implications of the rise of GAI in higher education and incorporate these new technologies effectively into their existing frameworks. The second, and arguably more significant challenge, is preparation: universities bear the responsibility of readying their students for a future of work that will be increasingly intertwined with automation.

Addressing these challenges demands a comprehensive exploration from both theoretical and empirical perspectives, which forms the essence of this paper.

Examining the impact of Generative AI on Asian top universities is crucial due to the distinct pedagogical traditions and the significant role these institutions play in global higher education. As these universities often serve as benchmarks for educational standards in the region, understanding how they adapt to and integrate GAI can provide valuable insights into the broader implications for Asian education systems. This examination can also highlight unique challenges and opportunities, shaping strategies for effective AI adoption that respect cultural nuances and pedagogical traditions in Asia. To a considerable extent, the quality of education in a country, particularly in the race for AI among major global powers, will also affect the level of national competitiveness in terms of both hard and soft power in the long run. Focusing on Asia provides a unique and diverse perspective to the discourse, given the rapid technological evolution of the region, its significant role in the global education landscape, and its potential as a leader in shaping the future of AI.

This research addresses several significant questions: How well have Asian universities integrated GAI into their policy frameworks? Are current curriculums prepared for the significant reforms that the advent of GAI might necessitate? What are the implications of these findings on the future of higher education and the readiness of our graduates for the future of work? The answers to these questions hold profound implications for the broader landscape of higher education, the future of work, education and human resources development policies, and national development.

In its research design and methodology, the paper is divided into two major and interconnected parts. First, it would conduct a thorough literature review to set up a theoretical framework about what learning and curriculum reforms would be needed to take up the challenge of higher education under the wave of Generative AI. Then, it would collect and analyze the documents on AI policy in teaching and learning issued by the top Asian universities. The top Asian universities as defined in this study would be universities located in Asia ranked among the top 100 by one of the two major rankings (QS and THE). In the latest rankings, there is a total of 25 Asian universities which would be included in our analysis. The AI policy documents of the universities will be coded and analyzed to address the research question of whether there is any gap between theory and practice in reforming the learning and curriculum of higher education for the era of AI. In addition to examining if major and necessary reforms are made to prepare students for the future of work under automation, it would also reflect on the mission and vision of universities in the digital age and the implications on education policy and outcomes. Furthermore, the influence of the AI perspectives, policies and curriculum reforms on the long-term competitiveness and evolution of the universities in specific and the Asian region in general would also be studied.

The journey of adopting GAI in higher education is fraught with complexities, and navigating this path requires deep introspection into these questions. By highlighting the need for this introspection, this paper aims to stimulate a much-needed dialogue and contribute to the broader discourse on the transformative role of GAI in higher education. It will start with the examination of the impact of AI on the future of work and its ability to reshape higher education in order to identify new capacities and curriculum changes needed before

surveying the official university policies to assess their preparation and readiness. Its findings will provide invaluable insights to educators, policymakers, and technologists, aiding them in their mission to harness the potential of AI while addressing the challenges it poses, thus shaping a future where human and artificial intelligence can coexist and thrive.

Governance of Health Data for AI Innovation

AI innovation in healthcare and state platforms under a rights-based perspective: the case of Brazillian RNDS

M. Matheus Zuliane **Falcão**[1], M. Raquel Requena **Rachid**[2] and Marcelo **Fornazin**[2]

1. *Centre for Law, Technology and Society, University of Ottawa, Ottawa, Canada*
2. *Oswaldo Cruz Foundation – Fiocruz, Rio de Janeiro, Brazil*

Sub. No: DAP-2023-0176 – Sp7
Full paper is accepted for publication in the Data & Policy journal.

Abstract

This article examines the National Health Data Network (RNDS), the platform launched by the Ministry of Health in Brazil as the primary tool for its Digital Health Strategy 2020 – 2028, including innovation aspects. The analysis is made through two distinct frameworks: Right to health and personal data protection in Brazil. The first approach is rooted in the legal framework shaped by Brazil's trajectory on health since 1988, marked by the formal acknowledgment of the Right to health and the establishment of the Unified Health System (SUS), Brazil's universal access health system, encompassing public healthcare and public health actions. The second approach stems from the repercussions of the General Data Protection Law (LGPD), enacted in 2018 and the inclusion of Right to personal data protection in Brazilian's Constitution. This legislation, akin to the EU's GDPR, addressed the gap in personal data protection in Brazil and established principles and rules for data processing. The article begins by explanting the two approaches, then it provides a brief history of health informatics policies in Brazil, leading to the current Digital Health Strategy and the RNDS. Subsequently, it delves into an analysis of the RNDS through the lenses of the two aforementioned approaches. In the final discussion sections, the article attempts to extract lessons from the analyses, particularly in light of ongoing discussions such as the secondary use of data for innovation in the context of different interpretations about innovation policies.

Signalling and rich trustworthiness in data-driven healthcare: an interdisciplinary approach

Jonathan R **Goodman**[1] and Richard **Milne**[2, 3]

1. *Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, UK*
2. *Kavli Centre for Ethics, Science, and the Public, Faculty of Education, University of Cambridge, UK*
3. *Wellcome Connecting Science, Hinxton, Cambridge, UK*

Sub. No: DAP-2023-0156 – Sp7
Full paper is accepted for publication in the Data & Policy journal.

Abstract

Discussions of the development and governance of data-driven systems have, of late, come to revolve around questions of trust and trustworthiness. However, the connections between them remain relatively understudied and, more importantly, the conditions under which the latter quality of trustworthiness might reliably lead to the placing of 'well-directed' trust. In this paper we argue that this challenge for the creation of 'rich' trustworthiness, which we term the Trustworthiness Recognition Problem, can usefully be approached as a problem of effective signalling, and suggest that its resolution can be informed by a multidisciplinary approach that relies on insights from economics and behavioural ecology. We suggest, overall, that the domain specificity inherent to the signalling theory paradigm offers an effective solution to the TRP, which we believe will be foundational to whether and how rapidly improving technologies are integrated in the healthcare space. We suggest that solving the TRP will not be possible without taking an interdisciplinary approach, and suggest further avenues of inquiry that we believe will be fruitful.

Operationalizing health data governance for AI innovation in low-resource government health systems - a practical implementation perspective from Zanzibar

Tracey **Li**[1], Abbas **Wandella**[1], Richard **Gomer**[2] and Mohamed Habib **Al-Mafazy**[3]

1. *D-tree, Zanzibar, Tanzania*
2. *School of Electronics and Computer Science, University of Southampton, UK*
3. *Information and Communications Technology Unit, Ministry of Health, Zanzibar*

Sub. No: DAP-2023-0157 – Sp7
Full paper is accepted for publication in the Data & Policy journal.

Abstract

Improved health data governance is urgently needed due to the increasing use of digital technologies that facilitate the collection of health data, and growing demand to use that data in artificial intelligence (AI) models that contribute to improving health outcomes. Whilst most of the discussion around health data governance is focused on policy and regulation, we present a practical perspective. We focus on the context of low-resource government health systems, using first-hand experience of the Zanzibar health system as a specific case study, and examine three aspects of data governance: informed consent, data access and security, and data quality. We discuss the barriers to obtaining meaningful informed consent, highlighting the need for more research to determine how to effectively communicate about data and AI, and to design effective consent processes. We then report on the process of introducing data access management and information security guidelines into the Zanzibar health system, demonstrating the gaps in capacity and resources that must be addressed during the implementation of a health data governance policy in a low-resource government system. Finally, we discuss the quality of service delivery data in low-resource health systems such as Zanzibar's, highlighting that a large quantity of data doesn't necessarily ensure its suitability for AI development. Poor data quality can be addressed to some extent through improved data governance, but the problem is inextricably linked to the weakness of a health system and therefore AI-quality data cannot be obtained through technological or data governance measures alone.

Assessing Health Equity in the IoT Era: A Study on Algorithmic Bias and Public Health Outcomes

Thokozani **Hanjahanja-Phiri**[1], Jasleen **Kaur**[1], Arlene **Oetomo**[1] and Plinio **Morita**[1]
1. *University of Waterloo*

Sub. No: 0840 – Sp7

Abstract

Introduction

Recent technological advances have led to the proliferation of NextGen data sources, encompassing data from various Internet of Things (IoT) devices, wearables, and mobile health applications. These advancements have revolutionized data collection and analysis, offering insights into health patterns and behaviours. The integration of IoT in public health has the potential to significantly impact social determinants of health and promote health equity. However, the accuracy and fairness of the data generated by these technologies need careful evaluation to avoid exacerbating existing health disparities.

The proposed study aims to evaluate health equity by examining the accuracy of algorithms developed for the Donate Your Data (DYD) initiative by ecobee, a company specializing in smart thermostats. Specifically, the study will assess the prevalence of false positives and false negatives produced by these algorithms, which is critical for understanding the reliability and fairness of IoT-derived health data. By doing so, the research intends to shed light on how IoT data can either mitigate or exacerbate health inequities across different demographic groups.

Methods

The study utilizes the real-time DYD dataset stored on the Google Cloud Platform (GCP), encompassing data from approximately 179,000 households collected between 2016 and 2021. This large dataset has been transferred to Microsoft Azure Gen2 storage as raw data for detailed analysis. The research will focus on identifying potential sources of bias and inequity within the algorithms used, which could inadvertently contribute to health disparities.

A comparative analysis will be conducted on data collected before and after the COVID-19 pandemic, incorporating demographic variables such as area of residence, dwelling characteristics, and household characteristics as proxies for socioeconomic status (SES). This approach aims to provide a nuanced understanding of health equity in different contexts, particularly in how health outcomes are influenced by technological disparities.

Results

The study's findings will quantify the prevalence of false positives and false negatives among different demographic groups. These will be categorized by device model, country,

province/state, city, and SES. An equity index will be created using predefined variables, including country, province, city, floor area, number of occupants, number of floors, age of home, and household size. This index will serve as a proxy for SES, allowing for a detailed analysis of the data's impact on health equity.

The prevalence odds ratio (POR) will be calculated as an indicator of the risk incidence of inequities. By comparing these ratios across various demographic categories, the study aims to provide critical insights into how the algorithms affect different populations. This analysis is crucial for identifying and addressing potential biases to prevent unequitable health outcomes.

Discussion

The study acknowledges the technical challenges inherent in IoT data, particularly the potential biases present in the data sources. It is essential to recognize that these biases may favour certain demographic groups, such as younger, more tech-savvy households with higher SES. This skew can lead to disparities in health data representation and outcomes, which need to be addressed to ensure equitable public health strategies.

By examining the accuracy and fairness of IoT-derived health data, the study aims to guide public health authorities in mitigating unintended effects such as bias and inequity. This effort is crucial for developing more inclusive public health strategies and policies, which can lead to a fairer healthcare system. The study's findings will provide valuable insights into the differential impacts of algorithms on diverse populations, informing future public health initiatives.

Therefore, this study represents a significant step towards understanding and promoting health equity in the context of IoT data. By identifying and addressing potential biases, it aims to contribute to the development of more equitable public health policies and strategies. The research highlights the importance of carefully evaluating and utilizing IoT data to ensure that technological advancements benefit all populations equitably and promote a more E-inclusive, or digitally inclusive, and fair healthcare system.

AI and Data Science to Strengthen Official Statistics

Measuring and reporting uncertainty of AI and machine learning tools in official statistics

Violeta **Calian**[1] and Anton Örn **Karlsson**[1]

1. *Statistics Iceland*

Sub. No: 3843 – Sp10

Abstract

1. Formulation of the research problem

In this paper we describe the Statistics Iceland's approach to uncertainty measurement and reporting when official statistics results are based on AI or machine learning algorithms. The main goal is to improve quality and reliability of the statistics publications while detecting, controlling and describing the limitations of this type of production processes.

We illustrate this proposal with two very different applications produced at Statistics Iceland:

(i) using Gaussian Processes (probability distributions over random functions which may also be mapped to less interpretable neural networks) for Bayesian forecasting of complex demographic data and

(ii) optimizing and evaluating classification algorithms for predicting the true resident population, for either census or survey design optimization

2. Methodology

The methodology follows the standard mathematical statistics approach to model fitting, model selection, out of sample prediction and uncertainty evaluation: we show that this general, scientific framework can be adapted and applied for the particular features of machine learning or AI models. Our solutions have several stages:

• exploring and describing the data (analysing distributions, correlations and clustering)

• training a set of algorithms, measuring their performance according to well defined metrics and identifying their optimum regimes based on the prediction goals (e.g. classification for census, survey optimization, or forecasting with generalised Bayesian models)

• reporting the uncertainty associated with the predictions. We emphasize this step as central to our study, since less frequently found in literature. We exemplify in the paper the quantification and reporting for:

(i) the uncertainty associated to the variability in the training data

(ii) the uncertainty due to the model fit and model complexity issues

(iii) errors due to distributional differences between training and predicting data sets

(iv) measurement errors

(v) errors due to interactions between epistemic (due to model) and aleatory (dominated by data) uncertainty

• describing the results in simple terms, by using interpretability tools which allow the user to understand the relations between the predictions and the features/variables involved in the AI/ML-model. We illustrate this stage with measures of feature importance, surrogate models in the case of complex classifiers or with conditional effects and posterior distribution checks in the case of Bayesian forecasting models.

3. Key findings

We conclude that both types of statistical products described on our paper, i.e. products based on new data science technologies and used for forecasting or classification purposes respectively, can be treated according to robust and transparent methods for measuring, controlling and reporting uncertainty. The only limitations to such a process may arise from insufficient computational resources, input data or incomplete domain/interpretation knowledge.

References

[1] Methodology of population projections based on Bayesian hierarchical models, Violeta Calian (2023) Working Paper in Statistical Series of Statistics Iceland, https://hagstofan.s3.amazonaws.com/media/public/2023/79a217c5-f567-4ddb-bed7-45329a32d531.pdf

[2] Machine learning estimation of the resident population, Violeta Calian, Margherita Zuppardo, Omar Hardarson (2023) Statistical Journal of IAOS , accepted for publication.

AI in German official statistics - from first steps to recent challenges

Florian **Dumpert**
Federal Statistical Office of Germany

Sub. No: 5483 – Sp10

Abstract

Official statistics face a variety of challenges worldwide. Driven by the increased possibilities of obtaining information and the progress in information technology, the demand for information from politics, the economy and society on the most diverse subject areas of official statistics is increasing. In order to meet this demand adequately, the production of statistics must be further developed. This is not only a matter of making new data sources usable, but also of making processes more efficient. This applies in particular to steps in the area of data processing (GSBPM phase 5), such as classify and code, review and validate, and edit and impute. In many NSOs (national statistical offices), solutions for (partial) automation of the processing steps are therefore being tested and used. Without the use of such statistical machine learning methods, it would not even be feasible to carry out some statistics due to their high frequency for instance. The talk will highlight classes of examples of how machine learning improves the production of German official statistics.

While (partial) automation aims at the efficiency of statistics production and possibly opens up new possibilities of data processing, the aspect of quality must not be neglected, especially in official statistics. Bad quality reduces trust very quickly. Existing frameworks at the level of the United Nations, at the supranational level (e.g. for the European Union) or at the national level frequently and rightly consider general requirements for the statistical institution, the processes and the statistical products. However, a concretisation for special situations, such as the use of machine learning, is necessary. There are first international and national works that address this concretisation. The talk will highlight the challenges and how to deal with them using conceptual and operational examples.

Harnessing Private Data for Public Policy: Organisational and Methodological Challenges, a focus on Mobile Phone and Card Transaction Data.

Marie-Pierre **Joubert**[1], Latifa **Oukhellou**[2], and David **Bounie**[3]

1. *INSEE (The National Institute of Statistics and Economic Studies)*
2. *COSYS-GRETTIA, Université Gustave Eiffel*
3. *Télécom Paris*

Sub. No: 4119 – Sp10

Abstract

Data from private origin represent a promising information source for National Statistical Offices (NSOs) and academic researchers. They complement traditional statistical sources by offering information with a higher temporal frequency or at a more refined spatial scale. Moreover, these sources have the potential to provide fresh perspectives on various issues, thereby enhancing the analytical capabilities of NSOs and researchers. However, using such data raises several challenges, encompassing access, interpretation, and processing. This article aims to shed light on these challenges through the prism of two distinct partnerships involving the French National Statistical Institute (INSEE), universities, and private entities. The article will review the historical developments in these collaborations, including the legal issues involved, and will also address the current working modalities and prospects for the future. The research subjects studied and the technical pitfalls involved in processing these new data, and integrating them with other data sources, will also be addressed.

The first partnership to be described is the MOBITIC project, an acronym for Mobility and Presence using Information and Communication Technologies. This project is financially supported until 2025 by the French National Research Agency (ANR) and involves collaboration between INSEE, the mobile operator Orange and the University Gustave Eiffel. The goal of this project is the fusion of various data sets, including mobile phone data aggregated at a detailed spatial scale, road vehicle counting loop data, public transport ticketing information, and comprehensive socio-demographic data collected by INSEE. The first research focus is day-time population (dynamic mapping of presence indicators at several spatial scales and several periods, zooming in on shopping town centres and on the identification of peripheral centralities). Then, using origin-destination matrices derived from mobile phone data and other data sources such as ticketing data and road counting data will enable to shed new light on official statistics territorial study zoning, for instance, functional urban areas. Indeed, territorial zoning rely mostly on census data related to commuting trips, overlooking factors such as telecommuting or journeys undertaken for

non-commuting purposes like leisure and shopping. These aspects could be better taken into account by incorporating insights from these new data sources. The research program extends its scope to the examination of dynamic socio-spatial segregation, encompassing changes in social mix over time and the isolation of disadvantaged areas. The integration of both novel and conventional data sources facilitates the generation of synthetic populations and will enable the simulation of multi-agent mobility. This simulation, in turn, will allow for an exploration of the implications of external events affecting transport demand, such as teleworking, sporting and cultural events, and anomalies in the multimodal transport network. Finally, the project aims to simulate the impact of public policy measures, including the development of teleworking, on congestion. The article will provide a more detailed description of these research issues and how fusion of conventional data with new data sources can address them.

The second project under consideration is INSEE's participation in the Digital Finance Chair. This research chair aims at conducting research using card transaction data to explore service, product and organisational innovations associated with digital technology, which have the potential to reshape the business of financial intermediaries. The primary objective is to use card transaction data for more accurate predictions of economic activity across various geographical and temporal scales. This chair involves collaboration with diverse partners. The data provider is the domestic card scheme, Groupement des Cartes Bancaires CB (CB), a private economic interest group that brings together a majority of France's financial institutions, with the goal of facilitating interbanking for payment cards. The scientific program of the chair is overseen by two academic institutions: Telecom Paris and University Paris 2 Panthéon Assas. Caisse des Dépôts et Consignations, a major public financial institution, and INSEE are the other financial partners. The data, anonymized for privacy, are accessible solely on CB's calculation servers and restricted to PhD students actively engaged in the project. The scientific program of the chair shares similarities with the research program of MOBITIC. Card transaction data, enabling a broad estimation of a cardholder's geographical trajectory through payment's locations tracking, will be employed for a comparative analysis with official statistics territorial study zoning. This comparison, along with results obtained on this theme from mobile phone data, promises to contribute significantly to the existing literature on the subject. Another major research focus will be on predicting retail activity, examining patterns of store visitation, and studying consumption externalities between stores. The chair also aims to investigate the economic impact of e-commerce.

Beyond the scientific content and governance structures of the projects, the article will highlight the practical aspects of the collaboration, underlining the differences between the two types of partnership. Given that these new data sources were not originally designed for official statistics, it is indeed essential to involve experts who are familiar with these data. These experts, members of partner private companies, possess a deep understanding of the data's architecture and the methodological nuances that demand careful consideration. Their contribution is therefore of great importance for carrying out high-quality analyses.

Advancing Public Diplomacy evaluations: AI and predictive analytics to leverage the global power of Hallyu, The Korean Wave

Natalia **Grincheva**
The University of Melbourne and LASALLE College of the Arts, University of the Arts Singapore

Sub. No: 9614 – Sp10 *(Remote Presentation)*

Abstract

My presentation will demonstrate preliminary results and reflect on the research project ***Mapping the Global Impacts of Hallyu, The Korean Wave*** (Read more: https://datatopower.net/hallyu). It aims to advance public diplomacy evaluations and measurements scholarship by focusing on the case of Hallyu (Korean Wave) that refers to the current global spread and impact of creative industries products specific to South Korean popular culture on different parts of the world. The project employs data-driven and Artificial Intelligence (AI) approaches combined with traditional qualitative research to measure, map, and predict the spread, reach, intensity and impacts of the Korean Wave soft power across different geo-locations to design more informed, strategic, and evidence-based approaches to public diplomacy evaluations and proactive management.
Research Context: Since the late 1990s, South Korean creative sector emerged as a speedily developing industry of transnational popular culture production. In the past decades, due to the developments of new media technologies it spread all over the world beyond Asia and the Middle East, and currently has a significant presence in Europe, North and Latin America. The Korean government capitalizes on the Hallyu global phenomenon that helps promote the image of South Korea as a modern, sophisticated, and technologically advanced society, which has also led to increased tourism and investments. The country's increased cultural exports up to US$ 10 billion reaching more than 157 million fans globally in 2020-21.[1] Beyond creative industries, the Korean Wave has also had a significant impact on South Korea's public diplomacy as a soft power resource that helps to advance its foreign policy objectives in different parts of the world and in many cases provides a convenient platform for diplomatic events, strategic geopolitical negotiations, and alliance building.[2]
Research Problems: The geography[3] of the Korean Wave from Japan to Mexico and its

[1] Korean Foundation for International Cultural Exchange (KOFICE). 2020. Research and research data. https://kofice.or.kr/b20industry/b20_industry_00_list.asp?mnu_sub=20100
[2] Kim, Youna. 2021. *Soft Power of the Korean Wave: Parasite, BTS and Drama*. Abingdon, Oxon; New York: Routledge; Cicchelli, Vincenzo. 2021. *K-pop, soft power et culture globale: Surfing the Korean Wave*. S.L.: Palgrave Macmillan; Jin, Dal Yong. 2016. *New Korean Wave: Transnational Cultural Power in the Age of Social Media*. Urbana, Chicago: University of Illinois Press. Kim, Youna (Ed.) 2013. *The Korean Wave: Korean Media Go Global*. London: Routledge.
[3] Jin, Dal Yong; Yoon, Kyong and Wonjung Min. 2021. *Transnational Hallyu: The Globalization of Korean Digital and Popular Culture*. London; New York: Rowman & Littlefield; Marinescu, Valentina. 2016. *The Global Impact of South Korean Popular Culture: Hallyu Unbound*. Lanham: Lexington Books; Kuwahara, Yasue. 2016. *Korean Wave: Korean Popular Culture in*

periodization from 1990s "made for Korea" to 2000s "made in Korea" to 2020s "made by Korea"[1] was explored in different studies. They revealed dropdowns and spikes of Hallyu presence and impacts across time and space, while suggesting that due to unfolding process of globalization and increasing media contraflows from emerging economies, like in China and India, Hallyu might lose its grip in the Asian and world markets and find difficulties to further penetrate them meaningfully.[2] Furthermore, in the recent years the negative aspects of Hallyu success have brought about serious concerns related to the Anti-Hallyu Movement. It can engender negative soft power of the country, which can be transformed into value clashes or propaganda, a flip side of the coin of the formation of soft power, reconciliation, or leadership.[3] Considering these trends in fluctuation of the Korean Wave spread, reach, intensity, and sentiment, either negative or positive, across different geo-locations, this project aims to test innovative data-driven approaches combined with traditional qualitative research insights to provide a more comprehensive account of the Hallyu soft power. In the context of Hallyu studies, for example, the scholarly and government research has created a whole "dataverse" of various sets of data which exist separately across different reports, databases, maps and raw materials.[4] The project provides a platform for a comprehensive exploration of the Korean Wave's three decades of currently disintegrated data from the moment of its inception until today. It explores how data-driven and AI approaches could be meaningfully employed in public diplomacy research to better understand the phenomena of soft power. Specifically, it investigates to what extent accumulation of big data and machine learning could advance public diplomacy research to measure, map, and predict diplomatic impacts of cultural exports and their attraction power spillovers.

Research Design: This project draws on the most research developments of the Data To Power geo-visualization application, a dynamic mapping software that allows to measure, map and predict soft power impacts on the global scale.[5] The application was created in collaboration with the Digital Diplomacy Research Group from the University of Oxford in

Global Context. Palgrave Macmillan; Kim, Youna. 2013. *The Korean Wave Korean Media Go Global*. London New York Routledge.

[1] Cicchelli, Vincenzo and Sylvie Octobre.2021. *The Sociology of Hallyu Pop Culture: Surfing the Korean Wave.* Springer Nature*;* Kim, Bok-rae, 2015. Lee, Sangjoon. 2015. Song, Sooho. 2020. "The Evolution of the Korean Wave How Is the Third Generation Different from Previous Ones?" Korea Observer - Institute of Korean Studies 51 (1): 125–50; "Decade of Hallyu Scholarship: Toward a New Direction in Hallyu 2.0." In: Lee, Sangjoon, and Markus Nornes. 2015. *Hallyu 2.0 : The Korean Wave in the Age of Social Media*. Ann Arbor: University of Michigan Press. "Past, Present and Future of Hallyu (Korean Wave)" *American International Journal of Contemporary Research* 5(15): 154-160; Chua, Beng Huat, and Koichi Iwabuchi. *East Asian Pop Culture: Analysing the Korean Wave*. Hong Kong University Press.

[2] Jin, Dal Yong. 2016. *New Korean Wave: Transnational Cultural Power in the Age of Social Media*. Urbana, Chicago: University of Illinois Press.

[3] Kim, HwaJung. 2023. "An Analysis of South Korea's Civic Virtue Soft Power." in Chitty, N., Rawnsley G.D. (eds.), 314-325, *The Routledge of Handbook of Soft Power*. Second edition. New York: Routledge.

[4] Hallyu Data is collected by Korean Foundation for International Cultural Exchange, Korea Creative Content Agency, Korea Tourism Organization, The World Association for Hallyu Studies, K-Culture Story Research Institute, Korea Culture & Tourism Institute.

[5] Grincheva, Natalia. 2023. "Translating data into soft power." *Pursuit*. January 9; Grincheva, Natalia. 2022. "Beyond the scorecard diplomacy: From soft power rankings to deep mapping explorations." *Convergence*. 28(1): 70-91; Palgrave MacMillan, 397-419; Grincheva, Natalia. 2019. "The Form and Content of 'Digital Spatiality': Mapping Soft Power of DreamWorks Animation in Asia." *Asiascape: Digital Asia*, 6 (1): 58-83.

2021-22.[1] Data To Power employs multi-layered mapping empowered by GIS that allows for a focused integration of different types of data through their cartographic display on multiple layers to visualize and evaluate interrelationships, coexistence, and processes of complex phenomena predominantly by exposing and comparing different data across layers.[2] This mapping method offers a reliable tool to visualize a landscape of distributed data values across different countries or geographic areas to build an inductive platform for exploratory spatial analysis, leading to pattern recognition and modeling.[3] The component of geo-visualization enables an "integrated approach" in public diplomacy research, setting a robust platform to meaningfully combine different methods of soft power impacts explorations, including (1) assessments of cultural, social, and economic resources, (2) evaluations of outputs, (3) network analysis, and (4) measuring public perceptions.[4] Converging all methods together, a dynamic mapping helps to compare and contrast various data sets either separately or in correlation to each other, revealing spatial patterns, highlighting "black" holes or knowledge gaps and exposing geographic zones of unique interest which require further deeper explorations. For example, visualizing Hallyu impact geographies though physical audienceship, online viewership and fandom data in correlation with secondary variables across countries, from economic GDP to social demographic and to ethnic/cultural diversity index can expose new "glocal" variables which can inform public diplomacy research to help assess soft power potential opportunities and challenges in a particular locale.

AI-enabled predictive analysis: Employing a prediction model of linear regression, Data To Power draws on the subset of AI, such as the supervised machine learning algorithm, to analyze multiple data sets to predict outputs. It builds a mathematical model of a set of data that contains both the inputs and forecasted outputs that consist of training examples, refined, and upgraded each time when the actual outputs are received.[5] In this way, the app uses all previously accumulated data to identify a formula to forecast results, transforming a mere data analysis exercise into a data intelligence system. [6] Correlating multiple variables with foreign policy historical context, current agenda and future strategies of South Korea in relation to a specific country or region, such a mapping can help identify sustainable pathways for bilateral or regional public diplomacy or flag concern areas on the global map for conflict mitigation. Focusing on the analysis of Hallyu as a case example and expanding soft power evaluations' methodologies, the project offers a new platform for innovative academic research that can proactively generate knowledge to inform the practice of public diplomacy.

---

[1] Grincheva, N. (2024) "Digital Soft Power" In Aday, Sean. (Ed.) *Public Diplomacy Handbook*. Edward Elgar Publishing (in press).

[2] Anselin, Luc, Ibnu Syabri, and Youngihn Kho. 2006. "GeoDa: An Introduction to Spatial Data Analysis." Geographical Analysis 38(1): 5–22; Mu, Wangshu, and Daoqin Tong. 2019. "Choropleth Mapping with Uncertainty: A Maximum Likelihood–Based Classification Scheme." *Annals of the American Association of Geographers* 109 (5): 1493–1510.

[3] Cho, Wendy K. Tam, and James G. Gimpel. 2012. "Geographic Information Systems and the Spatial Dimensions of American Politics." *Annual Review of Political Science* 15 (1): 443–60.

[4] Grincheva, Natalia. 2018. "Mapping Museum Soft Power: Adding Geo-visualization to the Methodological Framework." *Digital Scholarship in the Humanities*, 34 (4): 730–751.

[5] Russell, Stuart and Peter Norvig. 2021. *Artificial Intelligence: A Modern Approach*. London: Pearson education limited.

[6] Grincheva, Natalia. 2022. "Making museum global impacts visible: Advancing digital public humanities from data aggregation to data intelligence." In Schwan, A. and Thomson, T. (Eds.) The *Palgrave Handbook of Digital and Public Humanities*.

Demonstrating the preliminary results of Hallyu mapping, my presentation will contribute to the *Data for Policy* Conference agenda reflecting on the current challenges and opportunities of the future of public diplomacy decision making with the help of AI. Specifically, it will demonstrate the power of machine learning and predictive analytics modeling to improve proactive management of foreign policy and international relations, while flagging existing risks and downsides in data driven policy analysis.

https://dataforpolicy.org