FABRICS

A. Comninos & M. Konzett

EMERGING AI READINESS

CONTENTS

Alex Comninos is an information and communications technology researcher and consultant from South Africa who has published research on various topics including human rights and the internet, internet governance, cybersecurity, intermediary liability, mobile banking, and African political economy. Alex is a member of the Association for Progressive Communications and has consulted and conducted research for the Open Technology Institute, Freedom House, the OSCE and the World Bank. Alex is a podcaster at CYBERnyama.net.

Martin Konzett is a principal software architect and engineer, designer, artist and lecturer from Austria with over thirty years of hands-on experience with code, and an over twenty year track record in successfully delivering software solutions with core competences in Industrial Software Engineering (SE) and Human Computer Interaction (HCI). Martin has a strong focus on project operations, project culture and peopleware. He is a founding member of the Austrian ICT4D chapter and has consulted governmental, intergovernmental and corporate bodies.

Joy Liddicoat is a human rights lawyer and researcher from New Zealand. Joy was a Commissioner for the Human Rights Commission of New Zealand from 2002 to 2010, is Vice President of InternetNZ, and served as Assistant Commissioner (for Policy and Operations) at the Office of the Privacy Commissioner of New Zealand from January 2015 to May 2018. Joy is currently a researcher at the University of Otago focusing on the human rights implications of Artificial Intelligence.

PREFACE

FABRICS is a valuable and timely contribution to unpacking the field of AI and human rights. In a unique collaboration, Alex Comninos and Martin Konzett emphasise that we have an opportunity to unpack the dystopian versus utopian discourse of Artificial Intelligence (AI) and to develop more nuanced, critical, and rights affirming responses as the technology develops. This book therefore provides a timely guide for both those who are new to the field of technology and public policy, and for those who for whom it is familiar territory. The authors provide helpful insights into human rights and AI, such as the new forms of harm that might arise, and consider options for regulatory responses in a fast changing environment. The authors introduce and then critique new ideas and concepts for assessing human rights standards in AI, such as the right to an explanation, and helpfully distinguish between two interpretations of this right under the European General Data Protection Regulation. The authors consider algorithmic transparency, helpfully outlining why the concept of algorithmic transparency is too narrow, and propose wider system transparency in order to better understand algorithmic decision-making.

In this rapidly changing world, critical voices are vital if we are to navigate the regulatory tensions between fostering an enabling technological development and ensuring the promotion and protection of fundamental human rights. FABRICS is therefore a timely and helpful contribution to both the empowering prospects and complex challenges of human rights and AI.

Joy Liddicoat, Researcher, University of Otago

REGULATORY CONSIDERATIONS IN ARTIFICIAL INTELLIGENCE

A lot of industries seem to be overwhelmed by the rise of Artificial Intelligence (AI). When thinking about the future implementations of AI, some proponents dream of utopias while some opponents have dystopic nightmares. There is a lack of consensus in interpretations of the current legal and regulatory environments covering the implementation of computer systems leveraging AI and the related concepts of machine learning and deep learning. Interpreting, implementing and complying with the regulatory environment is not possible without unpacking the foundations of automated decision-making in general, and seeing AI at the moment as a tool to enhance systems rather than a living and autonomous subsystem in itself.

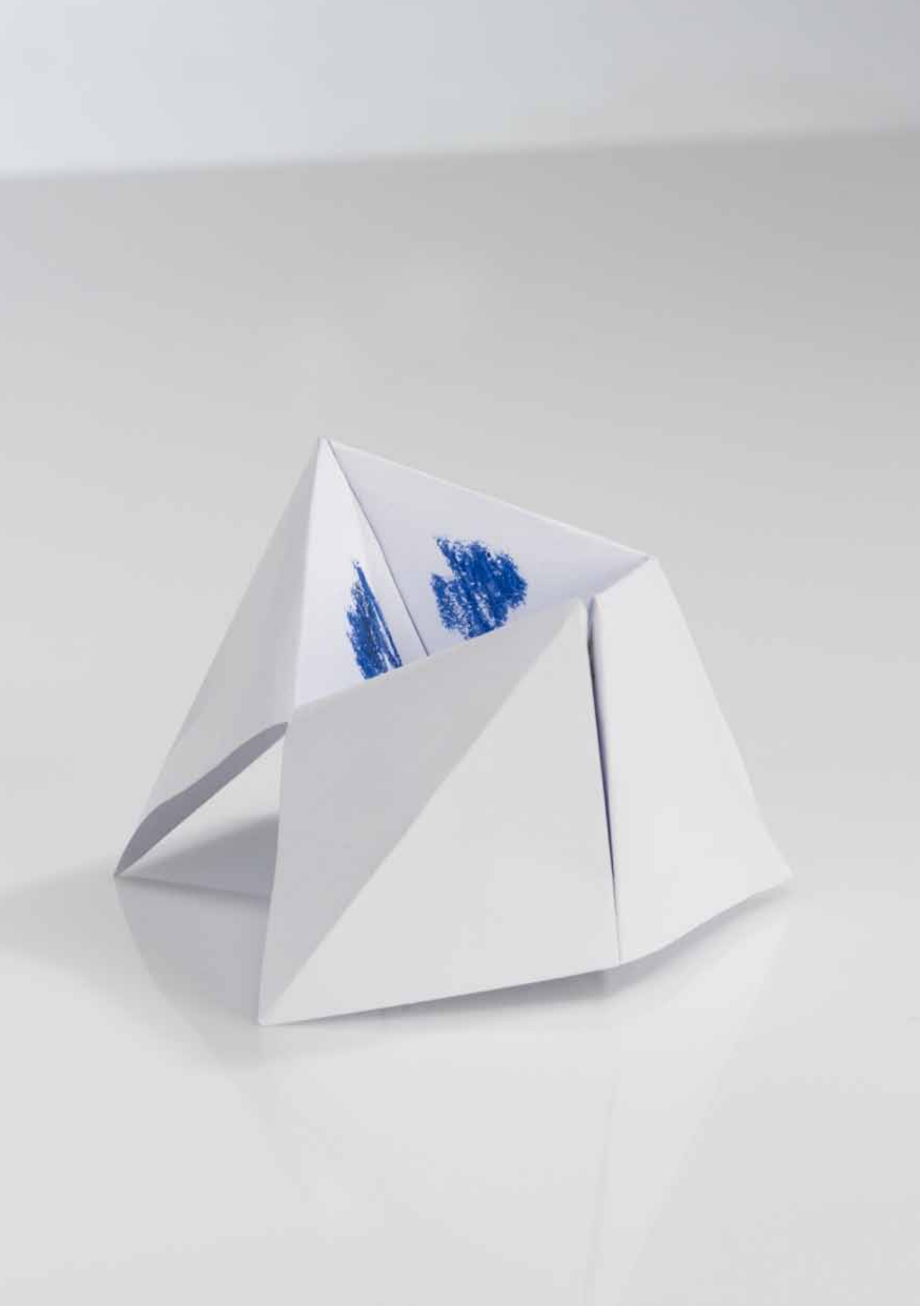
With the coming into force of the European Union's General Data Protection Regulation (GDPR) [1], as well as the adoption of data protection and privacy laws in over a hundred jurisdictions [2], regulatory and compliance issues with particular focus on privacy and data protection are now an unavoidable matter of fact for business. The GDPR has also introduced particular regulatory challenges to the use of AI in decision-making and the processing of personal data. The GDPR provides individuals, or "data subjects", with a set of rights regarding the use and processing of their personal data. Important to the implementation of AI are two following rights of data subjects that relate to automated decision-making.

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

https://eur-lex.europa.eu/eli/reg/2016/679/oj

[2] David Banisar, National Comprehensive Data Protection/Privacy Laws and Bills 2018, Updated 25 January 2018

https://papers.ssrn.com/abstract=1951416



Under Article 22 of the GDPR a data subject has the right to not be subject to a decision based solely on automated processing, including profiling, which produces legal or significant effects for them (except under certain exceptions). Data subjects also have a so-called "right to explanation" when they are affected legally or significantly by automated decision-making. While not expressed verbatim in the GDPR as a "right to explanation" - the extent to which this right exists, and how encompassing it is, is hotly debated - it can be interpreted as following from the text and principles of the GDPR. [3]

Some interpretations of the GDPR argue that these above restrictions do not apply in most cases, and that prohibitions on automated decision-making can be circumvented by employing measures like human intermediaries, acquiring consent from data subjects, and the use of contracts. This may be a tempting path for innovators and "disruptors", but is a short-sighted and perhaps a costly option. It banks on an assurance that regulators and courts will adopt the same interpretation. Should this not be the case, and stricter interpretations are adopted, there could be high compliance and liability costs in the future. Taking the GDPR seriously when it comes to automated decision-making creates opportunities to develop more understandable and auditable systems; leading to better insights from AI as well as systems that are easier to scale and augment.

Similarly, the GDPR could be an opportunity to focus on privacy by design, and on building secure products that are created to protect personal information. In light of the challenges posed by unpacking and implementing the rights of data subjects, all stakeholders involved - including users, engineers, systems administrators, product teams, corporations, and civil society organisations - need to come together to discuss a way forward. The following investigation shall provide some key regulatory considerations facing systems utilising AI and the related tools and concepts of machine learning and deep learning. One of the indicators of the current exponential growth in AI adoption, is as Gerard Verweij, global data and analytics leader at PricewaterhouseCoopers states, the increasing "consumer readiness to consume AI in all of its forms" [4].

At the time of the writing, almost all relevant software systems rolled out and used by both businesses and governments are already enhanced by AI, or have AI integrated into foundational components. This adoption of AI will soon require a human workforce in itself to implement and manage AI. This investigation also aims to provide insights for the workforce contributing to the design and implementation of such computerised systems within the context of the increasing relevance and usage of AI; and thus contribute towards deepening consumer readiness for AI into consumer preparedness to meaningfully and usefully adopt it.

SYSTEMS, CODE AND LAW

When talking about AI there is a tendency to talk in terms of utopias and dystopias. At one extreme is a techno-utopian dream once shared by an overlapping group of hippies and technologists in Silicon Valley that computing will give rise to a better world [5]. At the other extreme is a nightmare scenario where AI could become an existential threat. Between these two visions lies a more sober and perhaps boring truth; computers and systems, only do what they have been programmed to do. Extending this to the future of AI, it can be said that systems of Artificial Intelligence and learning in general, are equipped to make decisions, that are dependent on the rules that they have been provided, the data they have been provided, and the quality of that data.

The systems, no matter in which mode they are in performing tasks (e.g. learning, training, inferencing, or executing "decisions"), treat such provided data as facts (mediated of course by the defined weightings of the data points). Since the idealised world in which humanity is "watched over by machines of loving grace" [6] cannot be ensured, society must, to the best of its abilities, make sure that systems are equipped to make decisions that are ethical. In addition to being equipped to make ethical decisions, systems must also not break the law. Although within computer systems it seems to be that "code is law" [7], such systems should not break human law.

[3] Andrew D. Selbst and Julia Powles. 2018. Meaningful Information and the Right to Explanation, *International Data Privacy Law 7(4)*

[4] "PwC's Verweij Says We Are at the Revolution of AI", Bloomberg Markets, 25 July 2017

https://www.bloomberg.com/news/videos/2017-07-25/pwc-s-verweij-says-we-are-at-the-revolution-of-ai-video

[5] Adam Curtis, All Watched Over by Machines of Loving Grace (BBC Two, 2011)

[6] Richard Brautigan, *All Watched Over by Machines of Loving Grace* (San Francisco: The Communications Company, 1967)

[7] Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York: Basic Books, 1999)

[8] Wikipedia contributors, "List of autonomous car fatalities", Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/List_of_autonomous_car_fatalities (accessed 11 July 2018)

THE EFFECTS OF AUTOMATED DECISION-MAKING

Applications and systems leveraging AI must be compliant with all relevant national legislation and regulation, as well as international law and agreements. A lack of compliance can cause regulatory or legal liability as well as infringe on citizens' privacy, or cause harm or prejudice. Compliance of "code" with law, however is no simple matter. AI and automated decision-making in itself, introduces a whole range of legal challenges not yet fully understood by the stakeholders involved. Although legal and regulatory approaches are brought in place to handle AI, they appear and act at much slower pace than the emergence of new technologies.

At first, the question arises how AI-driven systems or even simple (non "intelligent") automated decision-making can harm or significantly affect people. AI-driven systems with direct kinetic or bodily impact are the most immediately evident when a system makes a decision directly resulting in physical harm or death. The possibility of systems navigating a self-driving car making decisions that cause the loss of life is no longer one of science fiction. By mid-2018 three drivers and one pedestrian had lost their lives in an accident with an autonomous system engaged [8]. Systems without such kinetic impact can also directly harm or directly and significantly affect humans. Such as when they are part of the intentional or inadvertent acquisition or misuse of data without consent.

Automated decisions can also negatively affect people by making erroneous, arbitrary, or unfair decisions that unfairly disadvantage or prejudice people. Examples could include automated or partially automated decisions regarding school and university placement, decisions regarding job vacancies, healthcare decisions, insurance premiums or decisions about citizens' access to government services. Unfair or false automated decisions can also have an impact on law enforcement and the justice system. Though less immediately tangible than the kinetic impact of a vehicle, such decisions can harm people through very real effects on a persons rights and freedoms as well as their physical, emotional, and financial wellbeing.

The extent of AI directly causing the above impacts is mediated by both the technical and the human systems in which the technologies are embedded. A decision that affects someone could be a human decision mediated by AI. Automated decision-making can also be mediated by humans. A human input or human decision taken before, during, after, or on the basis of a decision made by an AI could be the cause of harm. When someone is harmed by an automated decision, the question arises; is the fault with the automated decision-making system, or is the fault a result of a human action? This can be hard to establish in many cases. Both human and computers could be at fault, or to differing degrees.

When it is the system's "fault" however, one has to question the utility of blaming computers systems as computer systems cannot be held accountable for harm in a meaningful way, arraigned in court, or punished. On a deeper level the technologies discussed here remain social artefacts - systems that are socially constructed and socially reproduced by humans. They are designed, operated, reproduce and maintained by humans embedded in a society. Humans that are in the end responsible for creating the almost logarithmicly increasing amount and complexity of code that society finds itself embedded in.

Looking at decisions themselves as unit of analysis is helpful, but also important are broader macroscopic systems that are in place for managing aspects of the society and the economy - for example systems that exist to optimise public health or track and or predict global trade. These systems are mostly, at the moment, "only" there to assist decisions, not to "make" them, and the further implementation of results and/or insights are still manual (human) decisions and practices. These human systems are there because AI lacks a high degree of intelligence, but also for humans to act as gatekeepers to the implementation of automated decisions. As AI progresses in its utility and ubiquity, the scope, operation, and capacity of a human layer on top of such systems is not guaranteed unless it can be developed as fast as the technical layer.

MADE IN JAPAN
CANON LENS EF 50mm 1:1.4

ON PRIVACY BY DESIGN

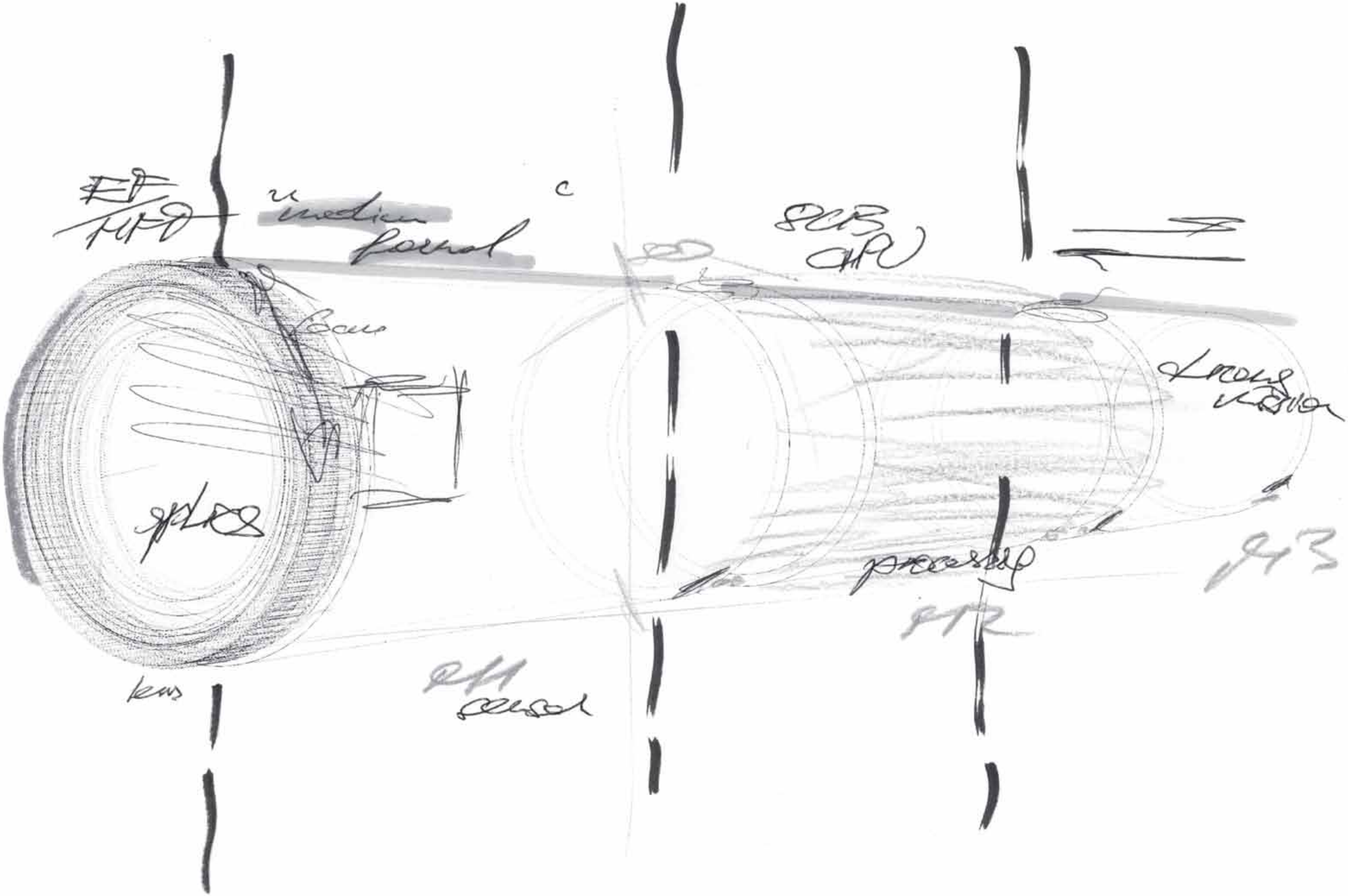
The practice of design fiction uses the design of fictional prototypes as a narrative device to explore and unpack possible futures. As science fiction author Bruce Sterling puts it, “design fiction doesn’t tell stories - instead, it designs prototypes that imply a changed world“ [i], thus helping to understand the near future in a critical way.

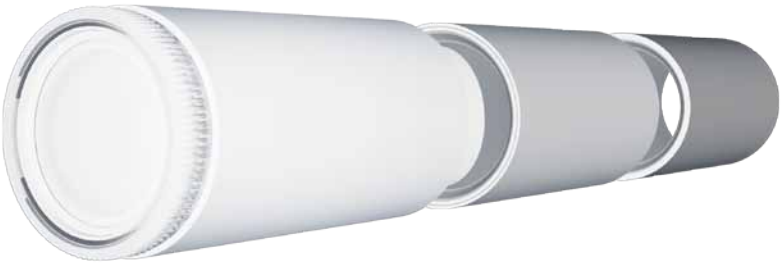
In *that new world* [ii], digital rights are enshrined in the constitution and actively protected by governance and management structures in the social and technological topoi. Data protection is a large and well developed area of such governance. The tech topos is engineered, produced and managed according to the principles of privacy by design [iii] - a fundamental principle in emerging information systems.

Privacy by design is entrenched in systems engineering and production, as well as in its deliverables. It is also entrenched in the social topos in engineering and corporate culture, societal institutions, as well as in legislation and regulation. The tech topos is designed, engineered, produced, and managed following the principles set out by Ann Cavoukian, Information and Privacy Commissioner of the Province of Ontario (serving from 1997 to 2014) [iv] and rules and methodologies introduced by pioneer software developer Alan Cooper [v] transponded into the (usability) engineering domain and formalised at best by Jakob Nielsen [vi].

Omega is a mature, senior-level, citizen in their mid-40s of *that new world* and is implemented in culture and infrastructure to build and protect trust and digital rights of citizens, users, and consumers as well as to empower them to exercise these rights. Ω is appointed to face the challenges of automated image capture and Artificial Intelligence (AI) based processing of such imagery. As an operator Ω is charged with upholding governance and management of citizen data and is an essential means of implementing the multistakeholder consensus, founded on industry claims that the future of a society leveraging AI needs “a workforce in itself“ as well as the need for “trust around data“ [vii].

Ω is charged with operation of an image capture and processing device that performs capture and processing tasks at the internet-of-things-based computing edge [viii] of the tech topos. The device is designed to be trusted to uphold data protection. Ω represents an archetype of a sentient user that operates this image processing software and hardware on the edge of the network, and before data transit to the fog or cloud. The considerations in architecting Ω as an operator, are the principle of least privilege, that access to personal data is kept to a minimum and only accessed when needed for pre-transfer processing, and that data is anonymised and stripped of personally identifiable information through preprocessing of data on the devices before the data departs to its destination. The device is radial and modular camera system to a) acquire image-based data sets for training of AI-based systems and b) to ingest imagery and inference based on that imagery on the edge of the tech topos (the periphery of cyberspace). The device showcases privacy by design by simply implementing the fundamental principle of visibility [ix] .

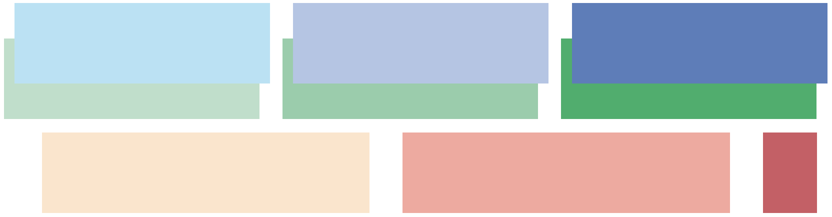

EF
MFD
u
medium
format
c
focus
ophRS
lens
sensor
SCB
CPU
Lrons
[illegible]
processor



Previous, Fig 1: Schematic drawing featuring Module 1, 2 and 3 (left to right). Module 1 is a Capture Sub-system with off-the-shelf optics (e.g. 30x45mm, 51.4-megapixel & CMOS sensor). Module 2 is an Inference Sub-system with an off-the-shelf GPU and NN-runtimes (e.g. SageMaker or Tensorflow). Module 3 is a Transmission, Sub-system for the encryption, rest and transit of data (e.g. over 5G).

Above, Fig 2: A monochromatic/grayscale application, coloured to be used in a urban context (exchangeable with different colour schemes).

Below, Fig 3: Colour schemes for a blue context (e.g. naval), a red context (e.g. arid) and a green context



[i] Bruce Sterling, “Patently untrue: fleshy defibrillators and synchronised baseball are changing the future”, *Wired Magazine*, October 2013

[ii] Michel Foucault, Of Other Spaces, 1967

[iii] Ann Cavoukian, *Privacy-Enhancing Technologies: The Path to Anonymity Volume 1* (Information and Privacy Commissioner of Ontario, Canada, 1995)

[iv] Ann Cavoukian, *Privacy by Design: Seven Foundational Principles* (Information and Privacy Commissioner of Ontario, Canada, 2011)

[v] Kim Goodwin. *Designing for the Digital Age: How to Create Human-Centered Products and Services* (Indianapolis: Wiley Publishing, 2009)

[vi] Jakob Nielsen. *Usability Engineering* (Morgan Kaufmann, 1993)

[vii] “PwC’s Verweij Says We Are at the Revolution of AI”, Bloomberg Markets, July 25 2017

[viii] Lopez et al. 2015. Edge-centric Computing: Vision and Challenges. *ACM SIGCOMM Computer Communication Review 45(5)*

[ix] Don Norman. *The Design of Everyday Things* (New York: Basic Books, 1988)

AUTOMATED DECISION-MAKING AND THE GDPR

As the macroscopic implementations of AI emerges, if the human layer cannot keep up, or becomes reduced, downsized, retrenched, or outsourced, the dystopic visions of AI will look like more credible threats. Governments and corporations are very tempted to invest in AI; such investments however need to occur in conjunction with investment in human capacity too.

The EU General Regulation on Data Protection (GDPR) came into force on the 25th of May 2018. Data Protection and privacy laws and regulations are not just a European development. There are over 100 countries, independent jurisdictions and territories that have adopted comprehensive data protection or privacy laws and there are 40 countries and jurisdictions with pending data protection bills or initiatives [2].

^

Decisions made by any automated means (from simple software systems up to AI - based systems) are subject to the GDPR when they process personal data or they have a significant effect on persons. Under Articles 14 of the GDPR, individuals or “data subjects” have the right to notification about the use of their personal data and the purposes for the use of their personal data. Under Article 15 data subjects have the right to access their personal data and be informed about how their personal data is used.

Under Article 22 of the GDPR, data subjects have the right “not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects [them]”. The EU Data Protection Working Party Guidelines on Automated Decision-making [9] interpret Article 22 as imposing a complete prohibition on fully automated decision-making including profiling when there is no human involvement in the process and when such decision-making has significant or legal effects [10]. Automated decision-making is not completely prohibited, but it is only permissible under certain exemptions. Automated decision-making can be allowed if:

i) the decision is necessary for entering into a contract,

ii) is authorised by the EU or an EU member states’ law,

iii) data subject has given explicit and informed consent for the use of their data.

[9] Article 29 Data Protection Working Party Guidelines on Automated decision-making and Profiling for the purposes of Regulation (EU) 2016/679

http://ec.europa.eu/newsroom/document.cfm?doc_id=47742

[10] The rather ambiguous “legal” effects mentioned in Article 22 of the GDPR are interpreted by the EU Data Protection Working Party Guidelines on Automated Decision-making as suggesting “a processing activity that has an impact on someone’s legal rights, such as the freedom to associate with others, vote in an election or take legal action” as well as a processing activity that affects the legal status of a person or their rights under a contract. With regard to the also ambiguous “significantly affects”, the guidelines argue that “for data processing to significantly affect someone the effects of the processing must be more than trivial and must be sufficiently great or important to be worthy of attention”, from the Guidelines cited by:
Sven Jacobs and Christoph Ritzer, Data Privacy: AI and the GDPR, Norton Rose Fulbright, 2 November 2017

https://www.aitech.law/blog/data-privacy-ai-and-the-gdpr

A RIGHT TO AN EXPLANATION?

Under the GDPR, data subjects have what has been called a "right to explanation" with regards to the reasons behind automated decisions that could significantly affect them. This "right to explanation", though not expressed verbatim in the GDPR as a "right to explanation" arises from a combination of rights afforded to data subjects under the GDPR [3].

A data subject has the right to be both given access to their own personal information, as well as to be informed about the use of their personal information. According to the interpretation of Norton Rose Fullbright, if automated decisions are made with personal data, the data subject needs to be informed that they are using their personal data in automated decision-making, "provide a meaningful explanation of the logic involved", and "explain the significance and envisaged consequences of the processing" [11].

Providing a meaningful explanation is easier said than done, the complexities and costs involved may incentivise companies to try exempt themselves from the need to provide an explanation.

THE TWO TYPES OF INTERPRETATIONS OF THE RIGHT TO EXLPAINATION UNDER THE GDPR

There are two sets of interpretations of the GDPR's right to explanation. The first is a strict interpretation and the second is a broad interpretation [12]. Under the strict interpretation, the right to explanation applies in a limited amount of cases, few decisions are based solely on automated processing, and usually a small amount of human intervention is included in decision-making processes. Furthermore, the vaguely defined significant or legal effects on the data subject are, in the restrictive interpretation, not actually produced. Proponents of the restrictive approach argue that it would be beneficial for data controllers - who would only have to disclose general information about profiling algorithms in limited cases [12,13].

The other, broader approach could be summed up as to play it safer with a wider interpretation of the prohibition on algorithmic decision-making. There may be more implementation challenges, but a broader interpretation more safely avoids legal and regulatory liability. Given the broad wording of articles 13, 14 and 22 of the GDPR. There is no guarantee that National Data Protection authorities or EU courts will also adopt the restrictive interpretation. It is argued by Malgieri and Commande that a broader "systemic open interpretation" might be more useful as it "enables a homogeneous approach tailored on the highest and strictest requirements, in order to be compliant with all possible interpretations at the national level thus reducing liability risks and compliance costs."

They also argue that "the duty to make algorithms legible" could help contribute towards opportunities for auditing software in white box scenarios (where the internal structure and components are known to the auditor). Such auditing could improve the quality of decision-making results and reduce liability risks [12].

Concepts of algorithmic bias, algorithmic transparency, and algorithmic justice, are often produced in a polemic manner, which can replicate the chasms between utopian and dystopian visions of AI. These concepts however do introduce important debates and questions regarding the design of future software systems or the extensions of the ones already in place.

Justice is a human system, comprised of human ideals. Software systems, while definitely (and not problematically) are able to affect the outcomes of the justice system, these systems cannot be just, only compliant. Impartiality, bias and transparency may be more useful concepts for understanding justice issues related to automated decision-making. AI needs to be equipped to make impartial decisions and harmful and discriminatory human biases should not be transferred to these systems. Engineers, designers and system architects must be conscious and mindful of their own biases in order not to transfer them to the machines.

In addition to biases held by individual humans, are biases that are constructed by the informational economy in which decisions are made; such as the accessibility and cost of datasets to train machine learning and deep learning systems. When systems make decisions that have a significant effect on others, we must be able to, when the request is reasonable and justifiable, explain to those affected the reasons for which an automatic decision was taken.

This is especially important when an individual is seeking recourse for the harmful effects of automated decisions. Such explanations, while important in avoiding legal and regulatory liability, will also be useful in future, adaptive systems or resulting databases, such that these mistakes are not replicated. They also may be important in the event that an automated decision is falsely implicated in a lawful decision, and thus important in defending those potentially legally liable for a decision in court. "Algorithmic transparency" is a not only misleading but also inaccurate terminology. A broader concept of system transparency needs to be proposed. System transparency involves the ability to explain why a piece of software, or in the best-case scenario, a well-defined software component, equipped with commonly known and open interfaces, AI-powered or not, has made a decision. System transparency should be an important objective framing the creation of automated systems but possibly may never perfectly be achieved in real world scenarios.

WARNING AND EMPOWERMENT

The International Organization for Standardization (ISO) has technical standards for graphical symbols conveying safety information on hazard and warning signs. Outlined in ISO 7010 [i], these standards convey information about prohibited or mandated activities, warn of hazards, and indicate safe conditions using colours and principles set out in ISO 3864 [ii].

Hazard and warning signs inform people around the world of safety risks in their environment in a universal manner that overcomes linguistic and cultural barriers. While “the consumer-readiness to consume AI in all of its forms” is there“ [iii], there is no accompanying universal semiotic framework for informing users of the potential threats and hazards posed to them by the presence of Artificial Intelligence (AI) in their digital and physical environment.

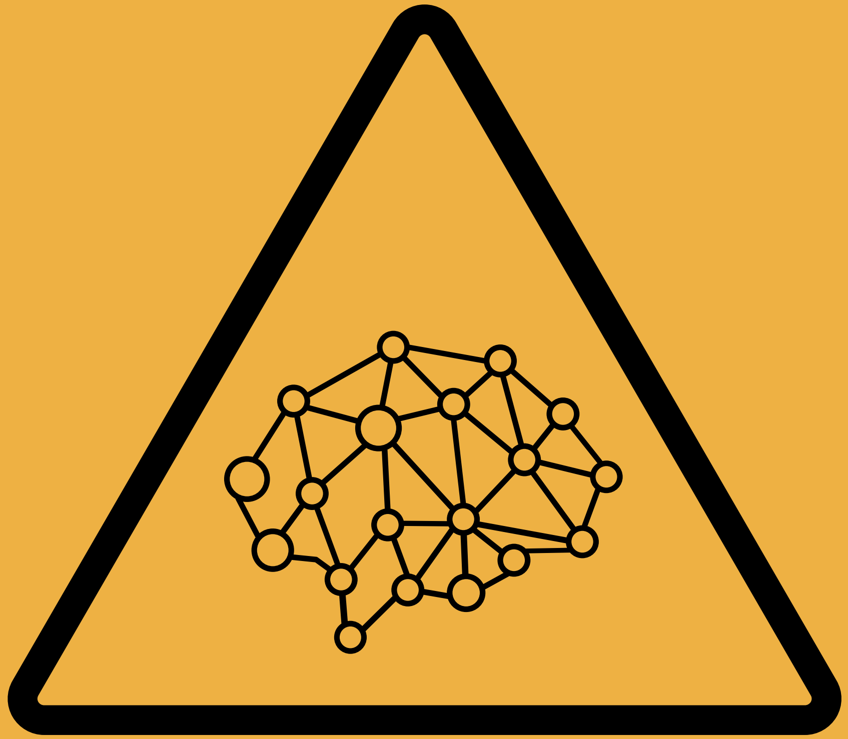
In light of the many social challenges accompanying the rise of AI, and that new world, as well as the need to deepen the exercise of the rights of data subjects set out in the EU General Data Protection Regulation (GDPR), there is the need to propose a framework of “Warning Signs for Consumer Empowerment“. The framework aims to move forward consumer readiness for AI and to deepen it towards consumer awareness, preparedness, and empowerment.

Below, Fig 1: Variant of a general warning sign that an AI-based system is in place. It serves to warn users that an automated decision-making system is in place. The sign warns that an AI system is in place and conveys information about the attributes of the system. It serves to warn users that an automated decision-making system is in place while conveying information about whether the system processes personal information, whether the algorithms can be explained by humans, and where the privacy policy of the system can be found.

[i] ISO 7010:2011. Graphical symbols - Safety colours and safety signs - Registered safety signs

[ii] ISO 3864:2011. Graphical symbols - Safety colours and safety signs

[iii] “PwC’s Verweij Says We Are at the Revolution of AI”, Bloomberg Markets, 25 July 2017



THIS ARTIFICIAL INTELLIGENCE

Processes personal information

□ YES □ NO

Is able to be explained

□ YES □ NO

Privacy policy at

____

Most software systems are executed in their runtime as "black box" – either obfuscated or proprietary and secret and thus illegible and unintelligible. Even if a piece of software operates in an open and transparent, "white box", the opacity of this box is relative. Code may not be organised and documented well enough to be understood by the entities in charge. As time passes and different programming languages, techniques and paradigms fade and new ones emerge into ubiquity, there are large amounts of code that are only legible to a decreasing pool of aging engineers.

Large and significant parts of legacy systems still operate on code bases created up to 50 years back, and written largely in now antiquated programming languages. The engineers able to maintain, compile code or even run the needed hardware is ageing. Such legacy systems are by no means only used for obscure and non critical purposes; they are quite prevalent in banking infrastructures as well as industrial control systems controlling critical infrastructures such as electricity grids. System transparency will come at a cost. As explanations of why automatic decisions are made become required in courts or to demonstrate compliance with regulations, the question of who will take on the costs of finding out why a decision was made will sooner or later become very important.

No matter how hard it is to build into a system the ability to assist in providing an explanation for a decision, or to build in transparency to assist in explanations of its behaviour, there is always cost involved. In an ideal world, for each component or layer established, a public database of metadata, holding the tree of pre-made decisions down to the one questioned, could be established as well. The integrity of this database could be insured as public, cryptographically ensured, distributed ledger. However, there is then running and maintenance cost to keep such a layer in place.

The same cost challenges would apply to achieving tranpsarency as a human/manual act of forensics such as debugging and reverse engineering – while the technology costs would be lower, these are expensive and time consuming activities. Engineers in most circumstances are incentivised to create systems for production as quickly as possible, rather than creating components that are legible, documented, and explainable. "Trial and error" was always a dominant and popular approach in software engineering, and is becoming more and more relevant with trends like serverless architecture and developing in the cloud with disposable infrastructures.

At the end of each project, the engineers moves on and code bases are often left behind, later picked up again, and so on. Within this context, the maintenance of such a lasting and meaningful system transparency will be hard to achieve. System transparency, though an important objective, will remain a Gordian Knot for some time. There are many layers and stacks of systems and thus of system transparency. Systems of automated decision-making include the systems themselves in production, the inherent features of a system, and the workforce and peopleware involved in designing, implementing, testing and running the system.

There are also third parties involved in systems engineering, delivering components to the systems in question, providing platforms to build on, or components and frameworks built on top of systems to complement or extend their functionality. Emerging legal and regulatory jurisdictions that attempt to govern and regulate automated decision-making, are beginning to influence systems and create new layers of peopleware.

There are multiple, overlapping, and sometimes conflicting, digital rights of citizens to balance and to ensure through different human rights instruments. All parties involved in creating systems or sourcing the datasets the systems uses to make decisions must be compliant with all applicable laws and regulations, as well as human rights obligations, at any point of the life cycle of a system.

[11] Sven Jacobs and Christoph Ritzer, Data Privacy: AI and the GDPR, Norton Rose Fulbright, 2 November 2017

https://www.aitech.law/blog/data-privacy-ai-and-the-gdpr

[12] Gianclaudio Malgieri and Giovanni Comandé. 2017. Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation. International Data Privacy Law 7(3)

[13] Sandra Wachter, Brent Mittelstadt and Chris Russell. 2018. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR, *Harvard Journal of Law & Technology* (forthcoming)

https://papers.ssrn.com/abstract=3063289

[14] Skala et al. 2015. Scalable Distributed Computing Hierarchy: Cloud, Fog and Dew Computing. *Open Journal of Cloud Computing 2(1)*

[15] Lopez et al. 2015. Edge-centric Computing: Vision and Challenges. *ACM SIGCOMM Computer Communication Review 45(5)5*

THE WAY FORWARD

There are two possible solutions to the regulatory challenges facing AI that could be implemented. An ethics and compliance tier comprising of a machine intelligible databases of legislation, regulations and international agreements in different jurisdictions could be strapped on to systems, either centralised and interfaced then by the system in question, or self-maintained as plug-in. Such tiers then need to be aware of the jurisdictional origin as well as the destination of data, and where it is stored and processed. Many data protection laws and regulations are based on the GDPR or the previous EU directive on data protection, and have similar principles and interpretations.

This may allow a simplification of the jurisdictional and legal instruction set and databases. In order to be compliant with the GDPR and other data protection laws, systems could be designed that minimise or fully eliminate personal information before processing. The more personal information (on data and meta-data level) is stripped out or anonymised through pre-processing at the place and time of capture (before transfer and further processing), the easier it is to avoid situations that need moderation. The rise of edge computing and fog computing [14] and the opportunities they present for processing data on devices at the edge [15] of cyberspace is thus an opportunity (as well as challenge and risk) to move forward in that direction.

This is where the discussion on* privacy by design *shall start.

FABRICS
Emerging AI Readiness
Alex Comninos and Martin Konzett

Impressum • Medieninhaber, Eigentümer und Verleger
Martin Konzett KG, A-4893 Zell am Moos, Österreich

First Edition, Zell am Moos, 2018
Published by Martin Konzett KG
pub@martinkonzett.com

ISBN 978-3-200-05921-4

https://comninos.org
https://martinkonzett.com

Idee, Konzept, Gestaltung und Produktion
Martin Konzett

Dank an Caroline Göllner
Reproduktion von *l'espace 1,2,3* Lukas Hämmerle

Alex Comninos and Martin Konzett are collaborating on *Emerging AI Readiness* as an ad hoc research unit under the working title VOUS. In the context of increasing industry and consumer demand for Artificial Intelligence (AI), VOUS aims to explore and unpack AI in all of its forms. Investigations can be found online, across social media, archived as GIT repositories, and are also available as paperback.

https://vous.ai
pub@martinkonzett.com

“In this rapidly changing world, critical voices are vital if we are to navigate the regulatory tensions between fostering an enabling technological development and ensuring the promotion and protection of fundamental human rights. FABRICS is therefore a timely and helpful contribution to both the empowering prospects and complex challenges of human rights and AI.“

Joy Liddicoat, Researcher, University of Otago

9 783200 059214