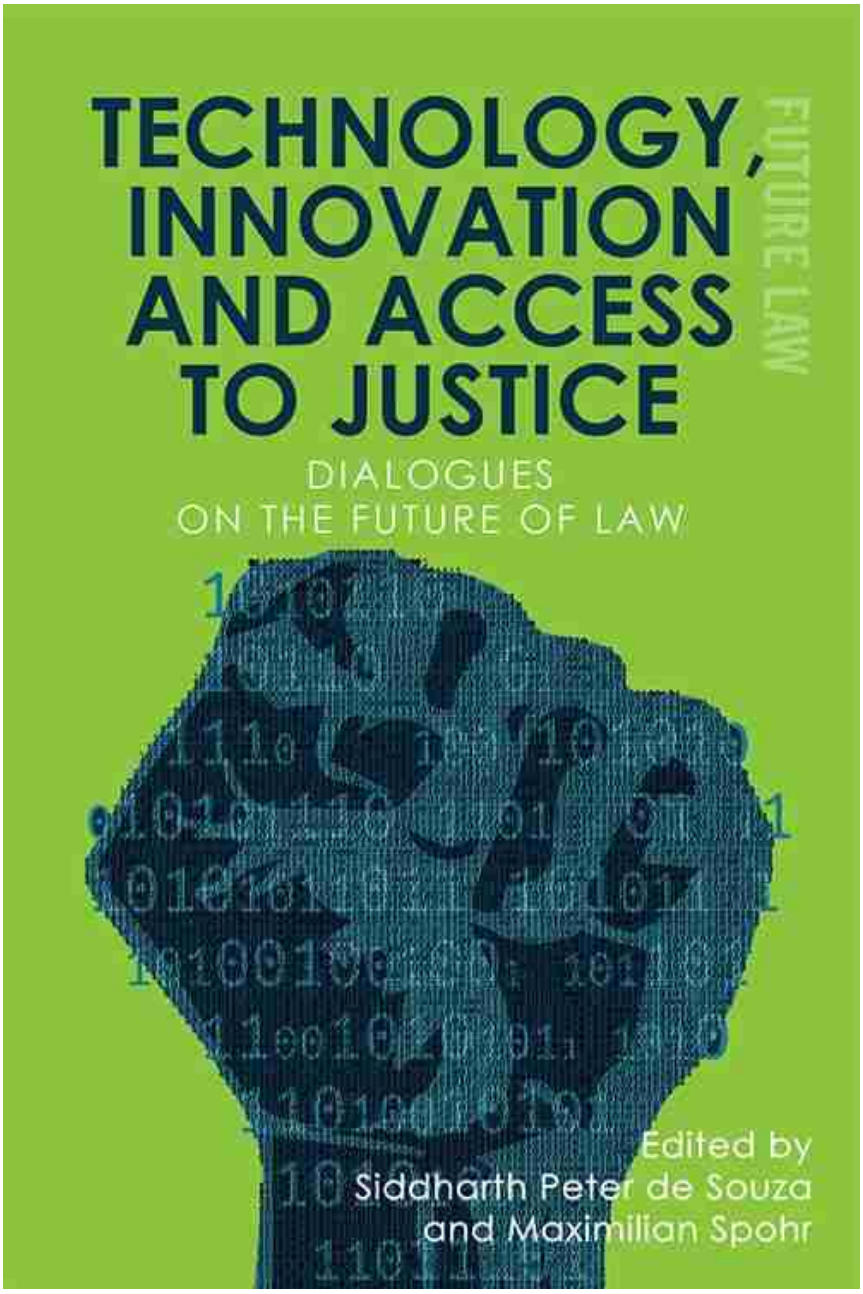


TECHNOLOGY, INNOVATION AND ACCESS TO JUSTICE
FUTURE LAW
DIALOGUES ON THE FUTURE OF LAW
Edited by Siddharth Peter de Souza and Maximilian Spohr

Djeffal, C. (2020). Sustainable AI Development (SAID): On the Road to More Access to Justice. In S. P. de Souza & M. Spohr (Eds.), *Technology, Innovation and Access to Justice: Dialogues on the Future of Law* (pp. 112–130). Edinburgh University Press.

7

Sustainable AI Development (SAID): On the Road to More Access to Justice

Christian Djeffal

A. Introduction

Artificial intelligence (AI) impacts society and our lives. Yet, there are very different ways to frame it. AI poses ethical, political, societal, organisational and economic questions. Scholars, politicians and other observers often use one of the frames to support or criticise AI. Fewer observers engage in the dis- cussion what the right frame should be and why we choose a specific frame. Therefore, this chapter looks into the potential of sustainable development as a frame for AI (Djeffal 2019b). Sustainable development is a framework that has not yet been in the centre of discussions surrounding AI, despite the fact that there is a huge potential to consider the transformative potential of digitisation and calls for a transformation for a sustainable future.

B. Sustainable AI Development

I. Artificial Intelligence

Artificial intelligence is a research question and area that is today dealt with by a whole subdiscipline of computer science. It aims to create intelligent systems, i.e. those which, according to Klaus Mainzer's working definition, can 'solve problems efficiently on their own'(Mainzer 2019: 3). Even the inventors of the computer had systems in mind that were supposed to per- form intelligent actions; one of their first projects could be described as a big data project for predicting the weather (Dyson 2014). The term artificial intelligence itself was coined by a group of computer scientists in a proposal to the Rockefeller Foundation to fund a seminar. They described their central research concern as follows:

> 'We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.' (McCarthy, et al. 1955)

In its origins, the concept of artificial intelligence was thus broad and reflected the intention to replace human intelligence with machines. Alan Turing foresaw that such projects would be criticised in his epochal essay 'Computing Machinery and Intelligence' (Turing 1950). In this essay, he dealt with the question of whether machines can think. His hypothesis was that humans will no longer be able to distinguish between human and machine intelligence after a certain point in time and that the question will thus lose relevance. So far, this has not happened; instead, two camps have formed. Some have pursued the so-called 'strong AI thesis', according to which AI can and will match and surpass human intelligence, while others, supporters of the 'weak AI thesis', have denied this and referred to the capacity of machines to solve certain problems rationally. This shows the fundamental disagreement in computer science about the goals and possibili- ties of this branch of research.

However, if the goals of the technologies are controversial, their develop- ment and eventual areas of application are likewise not predetermined. This is reflected in the dispute on whether AI should serve to automate human tasks or augment humans. This was already discussed in the early years of the AI debate (Grudin 2017: 99). One of the technologies that has brought artificial intelligence back on the map are so-called deep neuronal networks. These are adaptable non-linear mathematical models that are able to 'learn'. Since 2011, there have been several improvements that have led to an increasing hype around artificial intelligence.

II. The Impact of a General-purpose Technology

Like other technologies, one could describe AI as 'multistable'. This means that the scope and meaning of a technology in a society is only devel- oped over time and in the process of application and that these are not defined by the technology itself (Ihde 2012). What's more, AI is a general- purpose technology (Djeffal 2019a). By its nature, its purposes and its societal and individual consequences are contingent and dependent on its use.

Since AI technologies are flexible per se, they open up a new dimension of technical possibilities for action and reaction. It is not for nothing that the system is highlighted as an ‘agent’ from a computer science point of view (Poole and Mackworth 2011). As mentioned above, you could say that AI gives computers eyes, ears, arms and legs. Conversely, you could also say that cameras, microphones, loudspeakers and machines receive a brain.

If seeking to contrast AI with other fundamental innovations, one might meaningfully compare it with the ‘invention’ of iron. Iron is not a tool itself, but it is the basis for many different tools. Humans can forge swords or ploughshares from it. Iron also forms the basis for other technologies, be it the letterpress or steam engines. It is precisely for this reason that it is very difficult to speak in a general manner of the opportunities and risks of artificial intelligence. For what is seen as an opportunity and what as a risk often depends on how AI is specifically developed and used.

C. Sustainable Development as a Framework for AI

I. The Meaning of Sustainable Development

The notion of sustainable development was famously defined by the World Commission on Environment and Development (Brundtland Commission) as ‘development that meets the needs of the present without compromising the ability of future generations to meet their own needs’ (World Commission on Environment and Development 1987: 43). This is a vague but all- encompassing concept that forms the basis of a worldwide political process of setting goals and implementing them. Sustainable development has come to be a frame for agreeing on good policies in the international sphere. It is not in itself a goal but a meta-principle balancing several considerations in a specific way (Lowe 2001: 31). The concept of sustainable development is today universally referred to in international, national and local relations. The clearest expressions of that trend are the sustainable development goals of the United Nations. The Agenda 2030 describes seventeen goals the international community should work towards. The Agenda 2030 describes 169 targets in order to ensure the implementation of the goals.

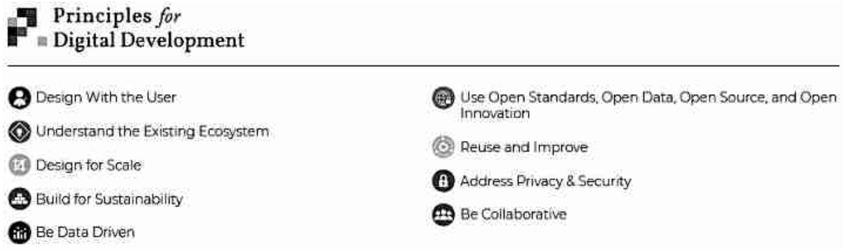
The concept of sustainable development was rooted in specific discourses in the 1970s and then proliferated substantially after that (Fukuda-Parr 2018). The discourse on sustainability was rooted in rising concerns about impacts on the environment. Environmental protection was one major driv- ing factor. This was due to the fact that the consequences of environmental harm became increasingly evident. Furthermore, it also became clear that certain natural resources were limited and potentially running out. Another important concern was the desire of less developed parts of the world to meet the needs of their populations and give them a good life. Decolonised states in particular called for justice and their right to development. They felt dis- advantaged. Environmental and development concerns both raised questions of justice and equity in different temporal regards. First, sustainable develop- ment relates to intragenerational justice, which means justice considerations between different people within one generation. Second, intergenerational justice concerns the justice between generations. This applies particularly to resources, but also to the behaviour of previous generations, particularly in the colonial context.

In a succession of activities at the United Nations, sustainable development became a major concept encompassing aspects and goals. Today, the goals laid down in the 2030 Agenda for Sustainable Development apply to different aspects transcending the categories of development and environment. The concerns included in the concept of sustainable development are represented by three pillars, namely the economic, the social and the environmental pillar. By tracing the different resolutions and declarations, the proliferation of the concept of sustainable development can be clearly mapped and we can see how it addressed an ever-increasing number of issues such as gender equality and access to justice. Sustainable Development Goal (SDG) 16 aims to 'promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels'. While commentators have criticised this aspect of sustainable development and stressed its contingency, sustain- able development has become a governance mechanism encompassing the whole world in a comparable manner. While the concept has no ascertainable fixed core, its content is fixed through a deliberative process.

This feature of being contingent but ascertainable could also be described as a major advantage. It conveys the understanding of a general process of balancing different considerations in a general conflict between maintaining and changing a status, between development and sustainability, between stasis and evolution. While different considerations might be relevant over a period of time, the principle of sustainable development can be a concept that offers room for all considerations if they are relevant for the general tension between sustainability and development.

II. The Case for SAID as an AI Framework

SAID is a very apt framework for designing, assessing and governing AI. The potential of the relation between artificial intelligence and sustainable development is sometimes touched upon but has not yet been fully explored.



Principles *for* Digital Development

- Design With the User
- Understand the Existing Ecosystem
- Design for Scale
- Build for Sustainability
- Be Data Driven
- Use Open Standards, Open Data, Open Source, and Open Innovation
- Reuse and Improve
- Address Privacy & Security
- Be Collaborative

Figure 7.1 Principles of digital development, derived from https://digital principles.org/principles/

In reflections upon AI, it is often stated that AI ought to be in compliance with human rights and ethical considerations. Commentators and policy makers stressed that stakeholders can use artificial intelligence to attain SDGs (United Nations Development Group 2017; Riegner 2016: 22; IEEE 2018). It is emphasised less frequently that AI applications can be used to further human rights and ethical principles (Djeffal 2018: 18). Development agen- cies have drawn up the ‘Principles for Digital Development’ in order to guide their development efforts.

The interesting aspect of these principles is that they directly relate to the assessment of technical artefacts. They provide for principles and criteria in order to make research and development sustainable (Principles for Digital Development 2017).

The same holds true for another framework published by the German Advisory Council on Global Change (WGBU – German Advisory Council on Global Change 2019). The report generally recommends using the digital transformation for a transformation towards sustainability. The report out- lines many ways in which digital technologies can be used to achieve the goals of sustainability. It also outlines general requirements for digital technologies that also apply to AI.

Linking sustainable development goals to AI development has several advantages. There seems to be a huge potential to understand and guide the process of research and development (R&D) of AI through the lens of sustainable development. The term development is part of sustainable development and of R&D. Furthermore, the underlying conflict addressed by the notion of sustainable development is also present in processes of introducing AI applications. Development is done in order to meet certain needs, but from a perspective of sustainability there ought to be limits due to a holistic view of other needs, be it of peers or of future generations. Most importantly, in the process of development, certain needs are in the foreground and sustainable solutions bring other less visible needs to the forefront. This general way of thinking (*Denkbewegung*) translates very well to the development of AI applications. They are often developed to fulfil certain tasks, whereas their unintended consequences and long-term impacts are not taken into consideration. This general fit might have its root causes in the fact that sustainable development was born out of the industrial revolution and is, therefore, translatable to what is today called the digital revolution.

Different considerations can play a part in this process of balancing the needs. While the 2030 Agenda for Sustainable Development addresses the digital divide in particular, other considerations such as data protection or cybersecurity could also be included in the framework of the SDGs. Its his- tory shows the adaptability of the principle of sustainable development. The knowledge and the questions that have been so far produced in the discourse around AI could very well also help to update the SDGs with new considera- tions and needs that will have importance for future generations.

One major advantage of using SAID as a framework is the inclusiveness of SD. The generic nature of sustainable development also leads to an all- encompassing design of the principle. The needs considered are not limited to certain categories such as human rights, societal interests or group rights. Sustainable development can account for all kinds of needs. This allows a more complete picture than would, for example, be garnered by focusing on human rights. Sustainable development is not a specific goal in itself, but a mission for continued awareness of the social, political and environmental consequences of our actions (Mulder, et al. 2011: 242). Another advantage is that environmental concerns have a self-standing value irrespective of whether they have immediate value to human beings. Sustainable development goals also set out positive goals that promise a better life on earth. Instead of framing questions in negative terms, such as discrimination or arbitrariness, sustainable development goals envisage positive ideals that should either not be impaired or furthered, such as equality and access to justice. These objec- tives are formulated in a manner that makes it possible to support them by specific measures and make progress visible by indicators. This verifiable and specific approach is very apt especially when it comes to processes of design. The specificity of the discourse and the general visions underlying them could help bridging communication gaps especially in multidisciplinary groups engaged in developing AI.

Sustainable development is always thought of in terms of the dimension of governance. This is expressed in the 17th SDG, addressing strengthening the means of implementation and setting out specific targets concerning gov- ernance. Reports and declarations on AI have seldom touched upon govern- ance issues and even more rarely on international governance mechanisms.

SDG 17 could have a very important impact here. This is also due to the fact that the governance of AI could be situated in existing fora that have an inclusive setting, allowing all states, along with different stakeholders, to enter the discussion. The governance mechanisms are well established and are constantly developed. They are forming on the international as well as on the national plane. This leads to a final decisive advantage of sustainable development as a framework for AI. SDGs are goals that are widely accepted in the international community. In many instances, they already provide guidance for AI development processes; in other instances, the development of the SDGs is necessary but also possible. It is very important to discuss the ethics of AI. Yet, an international agreement on the right international ethical standards on AI at the green table might be a long-lasting endeavour. In contrast, it is much easier to draw on work that has already been done in previous years and that resulted in actionable goals and an agenda that gives clearer guidance.

III. Layers of SAID

Commentators often denote sustainable development as a multi-dimensional concept. In order to arrive at a SAID framework, it will be important to address all applicable dimensions and implement the lessons the sustainable development process has learnt so far. Yet learning from AI discourses should also feed into the future of SDGs. Considering that AI-systems gain an ever- increasing importance, the future development of those systems might have a big impact on sustainable development. This would also serve as an active attempt to update the discussion on sustainable development, which dates from the industrial age, to the digital age, which is where we are heading. As a first attempt to map these dimensions, I propose three layers that should be addressed by SAID: the technology layer, the social layer and the governance layer.

a. Technology Layer

The technology layer translates issues of sustainable development to the level of specific applications. The goal here is to build 'sustainable technologies' (Mulder, et al. 2011). In line with this, technological design choices ought to be identified, highlighted and analysed. The important aspect is to link the design of technology to the goals pursued by SD. In that regard, SDGs can play different roles: the first role is to use technology for the realisation of the SDGs. In the case of AI, it would be to use different AI technologies in order to achieve sustainable development goals. The technology layer guides design choices. At the moment, there are several initiatives looking at what good design choices could mean. Examples pertaining to the technical layer relate to choices such as datasets and the architecture and design of algorithms. One issue is, for example, whether datasets to train algorithms are representative of all the people who will use the system later. If a speech recognition system is trained only with people with one particular accent, it will not understand people who speak differently. The choice of specific algorithms can make a big difference concerning their functionality but also their transparency. With deep neural networks with many layers, for example, the decision structures become increasingly blurred and harder to understand. This can have a big impact on access to justice, especially when the basis of a decision is blurred. These design choices have an effect on how the system operates and how it can be understood.

b. Social Layer

The social layer looks at the consequences of the use of AI systems in the social sphere. The focus of the sustainability analysis here looks at the impacts of the systems on individuals, groups and society as a whole. The dual effect of sustainable development also plays out in this field. On the one hand, AI can be used as a tool to achieve the goals set out in Agenda 2030. On the other hand, it should not be used to impair or contradict the SDGs. Regarding access to justice, there are applications facilitating citizens to make specific claims. Yet, microtargeting, for example, could aim at barring access to justice in certain situations. The social layer looks at the socio-technical reality of an AI system. This goes beyond a mere analysis of how the technol- ogy works. Especially the way in which AI systems are embedded in processes and the way their outputs are recognised socially are also looked into in this layer of analysis. In the case of algorithmic decision systems, one question would be whether there is a person able to exercise meaningful oversight over the system. Another question would be at which stage of a decision human oversight is possible or necessary. Therefore, the social layer looks at the design choices beyond technology. There are plenty of possibilities for how to embed technology socially. SAID is based on a focus and reflection process on how technology is embedded in real life.

c. Governance Layer

The governance layer looks at all the ways of influencing systems of artificial intelligence irrespective of the level (national, international, transnational). Together with sustainable development goals, a specific governance struc- ture was drawn up in order to realise the goals. As previously mentioned, SDG 17 addresses this very issue as it looks at and questions governance and implementation. In the discourse on artificial intelligence, govern- ance issues are raised on the national level (Tutt 2017) as well as on the international level (Gasser, et al. 2018). The governance level of SAID might complement the previous stocktaking and governance models in that it addresses specific governance challenges and problems more clearly. Naturally, it sees governance from a multi-level perspective, allowing for differences on the ground while stressing comparability between the differ- ent layers of governance.

d. Socio-technical Comparison

Together, the three layers give an account of what good design of artificial intelligence could mean. They are not to be perceived as existing in clinical isolation from each other but as different nodes in an active interaction. There is constant feedback and adjustment in order to improve design choices, social-technical settings and the right governance. If one looks at the layers as a choice architecture, one could compare it to a mixing console with three general areas and many ways to adjust the sound. There might be different possibilities to arrive at a good mix. Nevertheless, a requirement for a good mix is to understand the different layers and the attached choices.

Considering the many ways to achieve or not to achieve certain goals, a mere impact assessment will not be enough. If one approaches the implemen- tation of an AI system from a perspective of equity, the impact assessment can only be made if taking into account also the current state of affairs. The proposal made here is to make a socio-technical comparison that not only focuses on positive and negative impacts, but also assesses the current situa- tion (Djeffal 2018: 14). An analytical framework comprising different layers allows for socio-technical comparisons.

D. Use Cases Concerning Access to Justice

In the following, I will present two cases in which automated systems were developed over time and impacted on access to justice in different ways. The framework of sustainable development suggests looking at cases not only as being one point in time but also at their changes over time.

I. The ChatbotDoNotPay

In 2015 Londoner Joshua Browder programmed a chatbot *DoNotPay*. The idea was sparked when Browder received thirty unjustified parking tickets at the age of eighteen. He wondered how he could help people who wanted to take action against a parking ticket. He then successfully programmed a chat- bot which asked people simple questions in order to obtain the knowledge necessary for making their case.

a. Description

After an automated conversation, the bot advises people on the right course of action and potentially even returns a letter that they can use to send to their local authorities. In order to understand the administrative process and the relevant criteria, Browder filed several freedom of information requests. He programmed two versions for London and New York, which became a huge success. According to Tech Insider 3,000 people used the service, 250,000 parking tickets were appealed, with 160,000 successful appeals, saving the appellants a combined $ 4 million.

The chatbot was based on a decision tree and an automated document that resulted in a letter at the end of the procedure. With this relatively simple technical setup and sufficient background knowledge, Browder was able to help many people. The first development entailed a proliferation of the service. The chatbot was subsequently rolled out to all states of the US in 2017. What is even more remarkable is the fact that RoboLawyer was subsequently extended to ever more and more issue areas. With the chatbot having so much success, people contacted Browder to make him aware of other problems where a chatbot could be needed. This is when he discovered the problem of evictions and looming homelessness.

Collaborating with lawyers and several non-profit organisations, he went on to extend his chatbot to cover this topic as well. This new area revealed the limitations of such automation projects: while there was an enforceable right to housing in the UK, the situation in the US varied from one city to another. After the so-called Equifax scandal, in which sensitive data about many US citizens were stolen, a chatbot was created to help citizens to claim damages up to $25,000. Other chatbots were created to provide basic information and advice to asylum seekers, to help pursue charges because of unexplained bank charges or disputes between landlords and tenants. After further calls for help, Browder reconsidered his model of operation and implemented several changes. In order to assist people effectively, he aimed to include new issue areas. Therefore, he offered his technology as an infrastructure for people to develop their own chatbots. With this method, the RoboLawyer managed to extend to about 1,000 functions. They include the following:

- Sue anyone in small claims court for up to $25,000 without the help of a lawyer.
- Fight unfair bank, credit card and overdraft fees.
- Overturn your parking tickets.
- Claim hidden government and class action settlement money.
- Earn refunds from Uber and Lyft when a driver takes a wrong turn.

- Fix errors on your credit report.
- Save money on over 20,000 prescription and over the counter drugs.
- Scan your McDonalds, Jack In The Box, KFC and Carl's Junior receipts for free fast food.
- Find a California DMV Appointment in days rather than months.
- Apply for a United States B2 Tourist Visa extension or Family Based Green Card.
- Dispute fraudulent or low-quality transactions with your bank.
- Protect your privacy on Facebook, Twitter and Instagram. Sue big tech companies for every data breach.
- Make money on your airline and hotel bookings with price protection.
- Track your packages and earn refunds (or free Amazon Prime) for late package deliveries.

This decision had other impacts on the design of the chatbot. First, the chatbot was turned into a mobile app. After the increase of functions, the problem for users was to find the right chatbot. After gaining further funding for his project, Browder started to employ search machine technologies in order to allow finding the right service for people. Furthermore, the modes of interaction are to be improved through AI technologies. He is currently working on those functions after receiving substantial funds from investors and access to commercial AI technology.

b. Takeaways

This is a clear example of the employment of automation and AI technologies in order to further access to justice. DoNotPay aimed explicitly at helping people to claim their rights that would in other circumstances not be pursued by making it easier to produce legal documents or to have very easy guidance on how to act before a court. It is a good example of how technology can further an SDG. Another important aspect is how scaling such a technology is dependent on innovations and information on different levels. In the first phase, Browder understood that a problem he faced himself would be relevant to others. Later, he remained open for active input from people. They outlined their problems to him, and he took them up by including new functions. In this phase he worked together with legal experts, public administrations and others. In a third phase, he has provided others with technical building blocks to create their own chatbots. In a way, DoNotPay almost has become a technology platform for others.

If one were to look for the decisive innovation, many things had to do with social awareness and taking people's needs seriously. The underlying technology of the chatbot is comparatively simple. It was only in the later stages of the project that more refined AI technologies played a larger role. On the technical layer, it is interesting to see that the bot is able to have an impact without, at least for now, using state of the art machine learning tech- nologies. Despite the fact that everything was based on rule-based systems, Browder managed to scale the system. As regarding the social layer, the major point here is that he managed to be responsive to new problems and questions from the public. The RoboLawyer extended to areas in which it could have the greatest impacts. Looking at the perspective of a governance layer, it is significant how Browder managed to collaborate with public administration and with lawyers supporting his idea. It is a good example of civic technology, which is technology that is created by civil society to empower individuals in relation to civil society.

II. The Centrelink Issue

a. Description

Australia may be one of the countries which is most advanced when it comes to digitisation. Government and public administrations have been integral to this process. Notably, Australia managed to deal with a problematic applica- tion of digital administration in an open and transparent manner.

This concerned the so-called ‘online compliance intervention’ (OCI). The online compliance intervention consisted in further automating the processes of the Australian Tax Office (ATO) and taking civil servants out of the loop also to save costs (Belot 2017). Whereas previously only 20,000 cases a year were started, the projection was that automation would produce 783,000 interventions. It was aimed at raising and collecting debts, but ultimately resulted in a political scandal. An algorithm matched various tax- relevant data from two authorities. In order to achieve the right results, it used, among other things, fuzzy logic and techniques of data cleansing. The system found alleged contradictions between different data. Previously, those contradictions were taken up by civil servants dealing with cases. In a process of further automation, humans were taken out of the loop. After the changes, the system notified citizens automatically via SMS or letter about the alleged contradictions and asked them to make corrections in an online portal. If the citizens did not object, a payment notice was issued and the recipients had to object to it (Commonwealth Ombudsman 2017). This notice also contained a 10 per cent recovery fee as provided for byAustralian law. The algorithm looked for mismatches in the data. But the data is very error prone, mismatches can happen easily. The full automation meant that there were no civil servants involved in order to look for obvious or hidden errors before citizens were contacted. This, however, also limited the number of cases that could be dealt with. While the hope was to deal with more cases using less resources, there were many challenges to the system on different levels and a rising need for information. Because it was no longer possible to answer citizens' enquiries, temporary workers were hired and telephone contact with citizens was outsourced to a private call centre (Knaus 2017). People from weaker societal strata were particularly negatively affected as well as especially vulnerable or disadvantaged population groups, who could not defend themselves against the decision. The actual number of wrongfully issued notifications remains controversial.

b. Takeaways

Due to the open, thorough and transparent way Australian authorities dealt with the issue, it is possible to learn many things about how the implementa- tion of AI systems relates to access to justice on different levels.

On the governance level, it is a good example of a governance system that has enough watchdogs in place that can require improvements of the system and actually have done so. Firstly, Australian and international media engaged in the issue and highlighted problems continuously (Knaus 2016a; Martin 2017; McIlroy 2017; Medhora 2017; Towell 2017; Whyte). There were also watchdogs within the government that made suggestions to improve or to halt the system. One was the Commonwealth Ombudsman, an impartial and independent institution responding to complaints of the public and aiming at improving public service. The Commonwealth Ombudsman filed a report with many suggestions to improve the system (Commonwealth Ombudsman 2017). Another report was filed by a senate committee (Community Affairs References Committee 2017) What is striking in the case of both reports is that they deeply engaged with technological and social issues and made specific recommendations. Government replied to both reports and reacted specifically to the report of the Ombudsman (Australian Government 2017). While the Australian government was very well equipped for external review of such an incident, the same could not be said for its organisational govern- ance within the respective agencies. The entities responsible for explaining and reviewing decisions were obviously not well trained for the job. A media report suggests that critical employees even lost their job. Another general omission was that there was no random testing in order to improve the system. In comparison, the German tax administration is obliged to do some random testing according to § 88 of the German Fiscal Code.

The striking feature on the technical layer is that there have been hardly any technical changes to the matching algorithm. The same technology was used for the matching algorithm, but instead of making recommendations, the system decided automatically. While the affected citizens had the pos- sibility to correct data, it was the system issuing the final decision. A technical problem was that the system continued to make many assumptions that led to wrong results. The following mistakes were mentioned in one report:

- ‘income averaged over twenty-six fortnights in equal portions when the income was earned in a shorter time period;
- difference in employer’s name (for example, where a business name is provided to Centrelink and the ATO record includes company name) which resulted in the same income being duplicated; and
- non-assessable income considered assessable income such as a lump sum termination payment, paid parental leave and meal, laundry and uniform allowances’. (Community Affairs References Committee 2017)

Previously, these problems were compensated by humans overseeing the process and correcting some of the flaws of the system. Discrepancies were resolved through the human correction of the data or by contacting the respective employers. In the new iteration of the system, there was no feature of the system providing for an adequate quality. The system also relied on computing averages of certain values. This was sometimes favourable for the subjects of the decision, sometimes the debt was stated too high. Both reports criticised that there was no model that allowed to project what errors this averaging would produce. It was unclear what the relation of people with less and more money would be. With this feature, it would have been possible to change the process of averaging or to understand its risks better.

As regarding the social layer, it is interesting that there have been general problems of communication. The Ombudsman states:

> ‘Our investigation revealed DHS’ initial messaging to customers through its letters and in the system itself, was unclear and did not include crucial information such as a contact phone number for the DHS compliance team. Many complainants did not realise their income would be averaged across the employment period if they did not enter their income against each fortnight.’ (Commonwealth Ombudsman 2017: 2)

These communicative problems were in no way grounded in any feature of the technology, but in the face of the error rate of the system, they had huge effects. There was a lot of room for misunderstandings. Consequently, people missed the opportunity to correct their data. This is a good example of the facets of the transparency of a decision. It is far more than the ability to understand the basis of an automated decision. There has to be much more information. This communicative social aspect is all the more important if one considers the context. When considering social payments, the people affected are most likely not very skilled in accounting and correcting data. Therefore, there needs to be adequate communication for the social group to be addressed. One example for that would be to use 'Simple Language', that is an easy way to express which can be understood by people with different needs (Community Affairs References Committee 2017: 54ff).

One very interesting point about the discourse surrounding the OCI is the argument about what the error rate would be. The error rate of a hypothesis is defined 'as the proportion of mistakes it makes – the proportion of times that h(x) ≠ y for an (x,y) example' (Russell et al. 2016: 708). One disagreement during the discussions about the OCI was what an error would constitute. Some assumed that an initial notification by the system in cases in which there was only seemingly a debt due to incomplete or incorrect data was to be considered as an error of the system (Martin 2017; Pett and Crosier 2017). Others argued that in the initial stage, the system explicitly asked for more information and concluded that there was no error if the data was successfully corrected (Commonwealth Ombudsman 2017: 1). This disagreement relates to several questions such as who would be responsible to correct data in the system and how such a notice to correct the data is to be understood. Both views are viable, the difference maybe stems from a general shift in responsibility that is implicit in the further automation of the OCI. The further automation relied on the assumption that citizens are now solely responsible for their data management and public administration can merely rely on the data they provide. This shift in responsibility also means that citi- zens are ultimately responsible for correcting inconsistencies in their data. To ask them to correct the records in cases in which they have no debts is then not to be considered as an error. From the perspective of the citizens receiving the notice, this looks quite different. From their perspective, the hypothesis in the notice is that there is a debt and this is wrong (Knaus 2016b). A teacher receiving a letter because the system had mistakenly averaged his income in an incorrect manner, described his perception as follows: 'I was livid, absolutely livid. What really got me was the fact that the mentality behind it. It was like I was getting asked to pay back a loan.'(Whyte 2016). The fact that there has been such a profound misunderstanding and disagreement about questions of responsibility and error rates once again shows how important social factors are in setting up a technology.

E. Conclusion

Artificial intelligence is a general research question defining a field that deals with the independent solution of complex problems by machines. Under this umbrella, many technologies have been developed. When discussing these technologies, they are often framed in a particular way. We then talk about the ethics of AI or AI and human rights. Each frame highlights certain aspects and omits others. Therefore, I propose to aggregate different views about pursuing sustainability goals or designing digital technology in a sustainable way and to establish sustainable development as a frame for artificial intel- ligence. This allows to incorporate many learnings that have been made in the discourses surrounding sustainable development, but also to update these dis- courses and to include new learnings from the field of artificial intelligence. Another advantage is that the discourse around sustainable development is inclusive and pluralist while having an international range. Sustainable development also allows to analyse technology on different layers. On the technical layer, specific questions regarding the technology can be analysed. The social layer analyses the socio-technical surroundings of the technol- ogy that can be as important. Yet, sustainable development also looks at technological issues from a macro-perspective analysing the governance of the technology as a whole. The sustainable development goals have a double function. They are goals that should be supported by new technologies such as AI. They also give guidance on how AI should be developed generally and where developers should be careful. AI and access to justice are apt reflections of the aforementioned. The chatbot DoNotPay shows how technology can be used to give people access to justice. A simple chatbot allows people to formulate letters that they can send to public administration or to court. This system proliferated over a short period of time and now grants access in vastly different regards. The Australian Authorities used the online compliance intervention to collect debts from social benefits. In order to collect more debts, they turned an assistance system into a fully automated system. The burden to correct data was shifted from civil servants in public administration to citizens receiving social benefits. After several severe problems, two inquir- ies highlighted different issues with the system. Many of those remarks are hints on how to increase access to justice when drawing up a fully automated application with legal effects. An analysis structured by the different layers of SAID showed that many aspects have to be taken into consideration. These range from modelling the impacts of certain decisions and providing enough information on how to challenge the decision and who to ask for information to general governance questions about the organisation of publicadministra- tion. This example also shows that access to justice in the system of public administration has many faces. On the one hand, there is taxpayers' justice and the need for all people to pay taxes according to the same rules and not to receive unjustified benefits. But there is also access to justice when debts are reclaimed, and citizens challenge them. As in so many cases, justice is only achieved if there is an equilibrium of several views and needs. The concept of sustainable development explicitly addresses the question how to find such an equilibrium and convergence in the face of conflicts and opposing needs. This might be one aspect where sustainable AI development might turn out to be a frame that is useful for the analysis and design of artificial intelligence.

Bibliography

Australian Government (2017), *Australian Government response to the Community Affairs References Committee Report: Design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative*, available at <https://www.aph.gov.au/Parliamentary_ Business/ Committees/ Senate/ Community_ Affairs/ SocialWelfareSystem/ Government_Response> (last accessed 5 January 2018).

Belot, H. (2017), ‘Centrelink debt recovery: Government knew of potential problems with automated program’, *ABC News Australia*, 12 January 2017, available at <http://www.abc.net.au/news/2017-01-12/government-knew-of-potential- problems-with-centrelink-system/8177988> (last accessed 5 January 2018).

Commonwealth Ombudsman (2017), ‘Centrelink’s automated debt raising and recovery system’, available at <http://www.ombudsman.gov.au/data/assets/pdf_ file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery- system-April-2017.pdf>.

Community Affairs References Committee (2017), *Design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative*.

Djeffal, C. (2018), ‘Künstliche Intelligenz in der öffentlichen Verwaltung’, *Berichte des NEGZ*, 3, pp. 1–32.

——— (2019a), ‘Künstliche Intelligenz’, in T. Klenk, F. Nullmeier and G. Wewer (eds.), *Handbuch Verwaltungsdigitialisierung*, Wien: Springer.

——— (2019b), ‘Sustainable development of artificial intelligence (SAID)’, *Global Solutions Journal*, 4, pp. 186–92.

Dyson, G. (2014), *Turings Kathedrale*: *Die Ursprünge des digitalen Zeitalters*, Berlin: Propyläen.

Fukuda-Parr, S. (2018), ‘Sustainable Development Goals’, in T. G. Weiss and S. Daws (eds.), *The Oxford handbook on the United Nations*, Oxford: Oxford University Press, pp. 766–77.

Gasser, U., R. Budish and A. Ashar (2018), ‘Module on Setting the Stage for AI Governance: Interfaces, Infrastructures, and Institutions for Policymakers and Regulators’, *Artificial Intelligence (AI) for Development Series*, available at <https:// www.itu.int/en/ITU-D/Conferences/GSR/Documents/GSR2018/documents/ AISeries_GovernanceModule_GSR18.pdf> (last accessed 24 September 2018).

Grudin, J. (2017), *From tool to partner*: *The evolution of human-computer interaction,* London: Morgan & Claypool, 35.

IEEE (2018), ‘Ethically aligned design: Version 2’, available at <https://ethicsinac- tion.ieee.org/> (last accessed 2 June 2017).

Ihde, D. (2012), *Experimental phenomenologies*: *Multistabilities*, Albany: SUNY Press.

Knaus, C. (2016a), ‘Centrelink urged to stop collecting welfare debts after compliance system errors’, *The Guardian – International Edition*, 14 December 2016, available at <https://www.theguardian.com/australia-news/2016/dec/14/cen-

trelink-urged-to-stop-collecting-welfare-debts-after-compliance-system-errors> (last accessed 7 June 2019).

Knaus, C. (2016b), 'Government backs Centrelink debt system despite "incorrect" $24,000 demand', *The Guardian – International Edition*, 28 December 2016, available at <https://www.theguardian.com/australia-news/2016/dec/29/ government-confident-in-centrelink-debt-compliance-system-despite-reported- errors> (last accessed 5 January 2018).

Knaus, C. (2017), 'Centrelink to use 1,000 labour-hire staff to help recover welfare debts', *The Guardian – International Edition,* 21 November 2017, available at <https://www.theguardian.com/australia-news/2017/nov/22/centrelink-to-use- 1000-labour-hire-staff-to-help-recover-welfare-debts> (last accessed 5 January 2018).

Lowe, V. (2001), 'Sustainable Development and Unsustainable Arguments', in A. E. Boyle and D. A. Freestone (eds.), *International law and sustainable develop- ment: Past achievements and future challenges*, Oxford: Oxford Univ. Press, pp. 19–37.

Mainzer, K. (2019), *Künstliche Intelligenz – Wann übernehmen die Maschinen?,* Berlin, Heidelberg: Springer.

Martin, P. (2017), 'How the Centrelink debt debacle failure rate is much worse than we all thought', *The Sydney Morning Herald*, 25 January 2017, available at <http:// www.smh.com.au/federal-politics/political-opinion/how-the-centrelink-debt- debacle-failure-rate-is-much-worse-than-we-all-thought-20170124-gtxh8q. html> (last accessed 5 January 2017).

McCarthy, J., M. Minsky and C. Shannon (1955), 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence', available at <http://www- formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> (last accessed 31 March 2017).

McIlroy, T. (2017), '20,000 people sent Centrelink "robo-debt" notices found to owe less or nothing', *The Canberra Times*, 13 September 2017, available at <http:// www.canberratimes.com.au/national/public-service/20000-people-sent-cen- trelink-robodebt-notices-found-to-owe-less-or-nothing-20170912-gyg8mm. html> (last accessed 5 January 2018).

Medhora, S. (2017), '"This is a joke": fighting Centrelink's robo-debt, one year on', *ABC News Australia,* 22 December 2017, available at <http://www.abc.net.au/ triplej/programs/hack/centrelink-debt-one-year-on/9283108> (last accessed 5 January 2018).

Mulder, K., D. Ferrer and H. van Lente (eds.) (2011), *What is sustainable technology?*: *Perceptions, paradoxes and possibilities*, Sheffield: Greenleaf Publishing.

Pett, H. and C. Crosier (2017), 'We're all talking about the Centrelink debt controversy, but what is "robodebt" anyway?', *ABC News Australia,* 2 March 2017, available at <http://www.abc.net.au/news/2017-03-03/centrelink-debt-contro- versy-what-is-robodebt/8317764> (last accessed 5 January 2018).

Poole, D. L. and A. K. Mackworth (2011), *Artificial intelligence*: *Foundations of computational agents*, Cambridge: Cambridge University Press.

Principles for Digital Development (2017), 'Principles for Digital Development', available at <https://digitalprinciples.org/principles/>, (last accessed 7 June 2019).

Riegner, Michael (2016), 'Implementing the "Data Revolution" for the Post-2015

Sustainable Development Goals: Toward a Global Administrative Law of Information', *The World Bank Legal Review*, 7, pp. 17–41.

Russell, S. J., P. Norvig and E. Davis (2016), *Artificial intelligence*: *A modern approach*. Towell, N. (2017), 'Centrelink robo-debt: public servants removed for asking too many questions, says Andrew Wilkie', *The Canberra Times*, 16 January 2017, available at <http://www.canberratimes.com.au/national/public-service/centre- link-robodebt-public-servants-removed-for-asking-too-many-questions-says-andrew-wilkie-20170116-gts7na.html> (last accessed 5 January 2018).

Turing (1950), 'Computing Machinery and Intelligence', *Mind A Quarterly Review of Psychology and Philosophy*, 59, pp. 433–60.

Tutt, Andrew (2017), 'An FDA for Algorithms', *Administrative Law Review*, 69, pp. 83–123.

United Nations Development Group (2017), 'Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda', available at <https://undg.org/wp-content/uploads/2017/11/UNDG_BigData_final_web.pdf>, (last accessed 7 June 2019).

WGBU – German Advisory Council on Global Change (2019), *Towards our Common Digital Future*: *Summary*, Berlin: WGBU.

Whyte, S., 'How to dispute a Centrelink debt', *Crikey*, available at <https://www.crikey.com.au/2016/12/21/how-to-dispute-a-centrelink-debt/> (last accessed 5 January 2018).

Whyte, S. (2016), '"You feel powerless": Centrelink bullies are welfare collection cheats', *Crikey*, 13 December 2016, available at <https://www.crikey.com.au/2016/12/13/you-feel-powerless-centrelink-bullies-are-welfare-collection-cheats/> (last accessed 5 January 2018).

World Commission on Environment and Development (1987), *Our Common Future*, New York: United Nations, available at <www.un-documents.net/our- common-future.pdf>