



LUND UNIVERSITY

Human-Centred AI in the EU

Trustworthiness as a strategic priority in the European Member States

Larsson, Stefan; Ingram Bogusz, Claire; Andersson Schwarz, Jonas

2020

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Larsson, S., Ingram Bogusz, C., & Andersson Schwarz, J. (Eds.) (2020). *Human-Centred AI in the EU: Trustworthiness as a strategic priority in the European Member States*. European Liberal Forum asbl.

Total number of authors:

3

Creative Commons License:

CC BY-NC-ND

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Trustworthiness as a strategic priority in the European Member States

Edited by Stefan Larsson, Claire Ingram Bogusz
and Jonas Andersson Schwarz, with a commentary by Fredrik Heintz

AI Human-centred EU in the

ELF

The European Liberal Forum (ELF) is the official political foundation of the European Liberal Party, the ALDE Party. Together with 47 member organisations, we work all over Europe to bring new ideas into the political debate, to provide a platform for discussion, and to empower citizens to make their voices heard. ELF was founded in 2007 to strengthen the liberal and democrat movement in Europe. Our work is guided by liberal ideals and a belief in the principle of freedom. We stand for a future-oriented Europe that offers opportunities for every citizen. ELF is engaged on all political levels, from the local to the European. We bring together a diverse network of national foundations, think tanks and other experts. At the same time, we are also close to, but independent from, the ALDE Party and other Liberal actors in Europe. In this role, our forum serves as a space for an open and informed exchange of views between a wide range of different actors.

European Liberal Forum asbl
Rue d'Idalie 11-13, boîte 6, 1050 Brussels, Belgium
info@liberalforum.eu
www.liberalforum.eu

Fores

Fores – Forum for reforms, entrepreneurship and sustainability – is a green, liberal and independent think tank. It is dedicated to furthering entrepreneurship and sustainable development through liberal solutions to meet the challenges and possibilities brought on by globalization as well as the rapidly changing digital and increasingly data-driven society, and more. The principal activities of Fores are to initiate research projects and public debates that result in concrete reform proposals in relevant policy areas, such as on the digital development, including issues of AI and data, as well as on environmental policy, migration, entrepreneurship, and economic policy. We are a non-profit and independent foundation that act as a link between curious citizens, opinion makers, entrepreneurs, policymakers and researchers.

Human-Centred AI in the EU

Trustworthiness as a strategic priority
in the European Member States

Edited by Stefan Larsson, Claire Ingram Bogusz and Jonas Andersson Schwarz
with a commentary by Fredrik Heintz

Print: Spekter, Bulgaria 2020
Graphic design: Joakim Olsson (joakimolsson.se)

ISBN: 978-91-87379-81-9

This report is published by the European Liberal Forum asbl with the support of Fores. Co-funded by the European Parliament. Neither the European Parliament nor the European Liberal Forum asbl are responsible for the content of this publication, or for any use that may be made of it. The views expressed herein are those of the authors alone. These views do not necessarily reflect those of the European Parliament and/or the European Liberal Forum asbl.



Human-centred

AI

in the

EU

Trustworthiness as a
strategic priority in the
European Member States

Contents

	Commentary on AI in the EU	VI
1.	Trustworthy AI as a European Policy	14
2.	AI policy in Portugal Ambitious, yet laconic about legal routes towards trustworthy AI	48
3.	AI policy in Poland Ethical considerations already at the core	66
4.	AI policy in Norway Looking to the future and harmonised with the EU	86
5.	AI policy in the Nordics Pledging openness, transparency and trust, while expressing readiness to apply AI in society	106
6.	AI policy in the Netherlands More focus on practice than principles when it comes to trustworthiness	132
7.	AI policy in Italy Comprehensive focus on core infrastructural robustness and humanistic values	158
8.	AI policy in the Czech Republic Strong business focus, welcoming towards foreign investment	180

Acknowledgements

WE WOULD LIKE to extend our thanks to those who contributed to this volume; there is joy to be had in international collaboration, with diverse analysis and comments. We'd also like to thank the European Liberal Forum for the financial support and management that enabled the project. We'd like to thank Fores, particularly the head of the digital society program, Robin Vetter, and his exacting eye and quick laugh. We would like to acknowledge our peers at Movimento Liberal Social in Portugal who helped us plan for a Lisbon workshop that due to Covid-19 never materialised, as well as support from Institute for Politics and Society in Prague and Fondazione Luigi Einaudi in Rome. Last but not least: to Joakim Olsson for the adept editing and clean layout, to Ivan Panov for always-swift dialogue about print issues, and to Laetitia Tanqueray for her invaluable insights and nose for errors in grammar and text.

/ THE EDITORS

Commentary on AI in the EU

Fredrik Heintz

Fredrik Heintz, Associate Professor of Computer Science at Linköping University, Sweden. Coordinator of the TAILOR ICT-48 network, member of the European Commission High-Level Expert Group on AI (AI HLEG), director of the Graduate School for the Wallenberg AI, Autonomous Systems and Software Program (WASP), President of the Swedish AI Society, member of the CLAIRE extended core team, member of the EurAI board, and a researcher at the AI Sustainability Centre in Sweden.

THE EUROPEAN UNION has taken a clear stance on AI: we want AI, but we do not want any AI; we want AI that is Human-Centered and trustworthy. This means that AI is a means to improve life for us, not an end in itself. To be trustworthy, it has to satisfy the applicable rules and regulations, satisfy four ethical principles, and be safely and robustly implemented, as we in the High-Level Expert Group on AI have defined it. The Commission has started four networks of AI research excellence centres – the so-called ICT-48 networks after the name of the call – is running the AI4EU project,¹ and is intending to start a Public-Private-Partnership (PPP) on AI, data and robotics. In addition, and as focused in this highly valuable contribution, the Commission is also encouraging the member states to move in the same direction

¹ <https://www.ai4eu.eu/>

through the Coordinated Action Plan on AI. All of this is a great start! But, it is only a start.

This anthology provides an overview of the AI strategies of eight member states and Norway, and how they relate to the European priorities in AI, as outlined in the Ethics Guidelines for Trustworthy AI,² and more.³ As is clear, most member states are highly motivated with regards to leveraging AI, but they are approaching it from different perspectives. There is still significantly more work to be done to actually achieve the *orchestrated* effort that Europe needs. This volume serves the important purpose of displaying the intricacies and challenges, but also possibilities, of the European joint efforts of aspiring towards a value-based and trustworthy AI-development.

In this commentary, I focus on what I see as imperative for Europe to realise its vision of maximising the benefits while minimising the risks of AI in a coordinated European approach to human-centred trustworthy AI. With a steady emphasis on the educational needs linked to the AI-development, I cover definitional and regulatory concerns, as well as the importance of research and innovation. Concludingly, I envision a value-driven European AI-development at scale.

The AI definition: important, but moving target⁴

It is a major challenge to go from defining a research area to outlining a governance area.⁵ A key aspect of great importance is to get a reasonable working definition of AI that pushes the envelope rather than encompasses every digital system there is. AI is a moving target and will probably always be something we work towards, rather than something that is. Two important aspects are systems that do things that would require some cognitive functionality if done by people and systems that continually improve over time. These systems enable fundamentally new levels of automation and delegation.

2 AI HLEG (2019a).

3 See here also AI HLEG (2019b; 2020a; 2020b) as well as the main AI-policy documents from the European Commission, outlined in Chapter 1.

4 AI HLEG (2019c).

5 Larsson (2020).

Regulatory concerns

To me, the starting point is that humans should be held responsible and accountable. Why do we trust a pilot? Mainly because we believe that the pilot wants to survive as much as we do, and therefore will do everything in their power to land safely. The question then basically becomes: what is needed for us to trust a machine sufficiently to take the responsibility for the outcome of its actions? An AI system has no skin in the game and is therefore not really impacted by the results, nor punishable. In general, especially considering that we want these systems to complement us, I think we should strive to have the same requirements on AI systems as on people. We are the baseline upon which these systems should improve. The purpose is to get systems that raise the bar both in terms of capability and in quality compared to us.

Research

Europe cannot be a leader in AI regulation without being a leader in AI, and it cannot be a leader in AI applications or innovations without being a leader in foundational AI research. This necessitates a European research community that can unite through strong collaboration, and that can join forces with industry and society at large to build on European research strengths and enhance Europe's well-being. To achieve this, we need dedicated, significant and long-term research funding for both fundamental and purpose-driven research on AI to promote AI that is trustworthy and to address relevant scientific, ethical, sociocultural and industrial challenges. This is a necessary complement to the regulatory concerns.

Innovation

Europe has many small companies and startups, but very few of these scale, instead they have a tendency to get bought out by investors from outside Europe. It is therefore important to develop policy instruments that address this. The interaction between fundamental research and other functions in the innovation ecosystem needs to be substantially increased, and time from research to market needs to be shortened. To achieve this, it is necessary to establish a clear strategy for coordinating

and structuring an AI-based innovation ecosystem across Europe. There is a need to change existing policy instruments and strategies to take into account the significant role of entrepreneurs and private capital in the modern, AI-driven innovation economy. Europe currently does not create enough new businesses destined for growth and has relatively few innovation ecosystems of strength and coherence. European AI centres should therefore be established with the explicit mission of building and growing the European AI innovation ecosystem.

Education

The biggest challenge related to harnessing the power of AI is probably education. It is quite clear that the question is not about humans *or* AI, but rather how to best structure the relation between humans *and* AI. One important observation is that it is a different skill to, for example, play chess with a computer compared to playing chess without a computer. This means that even if you take the best expert in your organisation and give her the best tool, the result might not necessarily be better than before. A significant consequence of this is that we need to learn how to solve problems together with computers and we need to organise the work to support this new way of working. To me, the most important skill is computational thinking,⁶ which is all about solving problems using methods from computer science and using computers as tools.

Additionally, the amount of knowledge and data available to make decisions increases exponentially. If we assume that the amount of knowledge doubles every second year, then every two years you need to learn as much as you have learned since you were born, or base your decisions on less and less knowledge. This has major consequences for education. First, the amount of knowledge you need to learn after you finish your education is significantly larger than what you learned during your education, or you will fall behind. Second, the fraction of knowledge taught in school will be smaller and smaller. Besides improving and adapting education, we also need to use AI-based tools to deal

6 Wing (2006).

with this rapidly increasing amount of information and knowledge to make decisions.

Education is fundamental. Europe already has good educational systems that can be further improved. First, there is a need to significantly increase the volume of broad AI educational programmes with a focus on technology (at all levels including BSc, MSc, PhD, and post-doctoral). Second, develop specific AI educational programmes with a focus on dissemination in other sciences and society as a whole (again, at all levels including BSc, MSc, PhD, and postdoctoral). Third, make sure that primary and secondary education provides the necessary theoretical and practical foundations to allow everyone to become active and engaged citizens in the modern society, where AI is a natural part. We should also develop and implement a European Curriculum in AI to make it easier for individuals and companies to understand what knowledge is offered and expected.

To address more immediate needs, we also need to invest both in upskilling and reskilling people. There is a major need for AI talents with expert knowledge, who are capable of driving, managing and conducting AI activities in their institutions and organisations. Europe also needs to attract, develop and retain a comprehensive talent pool of AI developers, entrepreneurs and data analysts, and to create a beacon for talent.

The necessity of scale

Europe is doing many good, but relatively moderately sized, initiatives. To really make a difference and to take the next qualitative step, we need to significantly scale up these initiatives! Europe also needs an AI lighthouse, a CERN for AI, a single physical place with the attraction of the major AI hubs outside of Europe. The purpose is to effectively achieve critical mass, synergy, and cohesion across the European AI ecosystem without permanently dislocating talent from where it is needed the most. We need to make sure this is focussed on excellence and a site selection process grounded and transparently managed on the basis of politically neutral, externally validated criteria. The

lighthouse should be a symbol for European ambition and achievement in this area, a global magnet for talent, and the centrepiece of an AI ecosystem that spans all of Europe and all areas of AI. It should be “the place to be” when it comes to AI research and innovation in Europe. Somewhere people can meet for a period of time to work with other leading researchers and experts from all over the world on the most exciting and important topics, technologies and applications of AI. Through sabbatical and other temporary scientific positions, the hub will not drain talent from labs around Europe. Rather, it will act as the beating heart of European AI, a place where knowledge is mixed by the visiting researchers and then spread out again to the labs in the network with the returning researchers, thereby strengthening the development of excellent AI research across all of Europe.

AI for good and AI for all

Finally, focus “AI made in Europe” on “AI for Good” and “AI for All”. We should take global leadership in supporting publicly funded, large-scale AI research and innovation with a clear focus on the good of our citizens, our society and our planet. We should aim at creating intelligent machines that implement fundamental and shared values, respect and amplify human abilities and support the shaping of a better society. We should maximally leverage AI for achieving the UN Sustainable Development Goals – “AI made in Europe” should be “AI for Good”. It is also important to embrace the diversity of the different regions and cultures in Europe, making sure that the AI framework benefits all of Europe and leverages the talent and resources our diverse regions and societies have to offer. The European approach to AI should foster the accessibility of knowledge and broadly deployed technology by everyone, across different generations, with or without specialised education, by lowering the barrier to entry for the effective, safe and beneficial use of AI – “AI made in Europe” should be “AI for All”.

References

AI HLEG (2019a) *Ethics Guidelines for Trustworthy Artificial Intelligence*, Brussels: European Commission.

AI HLEG (2019b) *Policy and Investment Recommendations for Trustworthy Artificial Intelligence*, Brussels: European Commission.

AI HLEG (2019c) *A Definition of AI: Main Capabilities and Disciplines: Definition Developed for the Purpose of the AI HLEG's Deliverables*, Brussels: European Commission.

AI HLEG (2020a) *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment*.

AI HLEG (2020b) *Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence*. July 2020.

Larsson, S. (2020). On the Governance of Artificial Intelligence through Ethics Guidelines. *Asian Journal of Law and Society*, 1–23.

Wing, J. (2006). Computational Thinking. *Communications of the ACM* 49(3), 33–35.

CHAPTER 1.

Trustworthy AI as a European Policy

Stefan Larsson

Stefan Larsson is a lawyer (LLM), senior lecturer and Associate Professor in Technology and Social Change at the Department of Technology and Society at Lund University, Sweden. He holds a PhD in Sociology of Law as well as a PhD in Spatial Planning. His multidisciplinary research focuses on issues of trust and transparency on digital, data-driven markets, and the socio-legal impact of autonomous and AI-driven technologies.

Claire Ingram Bogusz

Claire Ingram Bogusz is a post-doctoral researcher at the House of Innovation at the Stockholm School of Economics and the Department of Applied IT at the University of Gothenburg, Sweden. Prior to completing her PhD, she read law (LLB). Her research interests lie in how code-based technologies change organising, professions, and the dynamics of value creation.

Jonas Andersson Schwarz

Jonas Andersson Schwarz is a senior lecturer and Associate Professor in Media and Communications at Södertörn University, Stockholm, Sweden. His primary research interest lies in the epistemological and ethical aspects of digital media infrastructure.

1. Introduction and purpose: AI and trust in Europe

DATA HAS COME to be seen as the new oil.¹ But, as with oil, it is not just control of the raw material that is valuable: being able to refine and process it into something more brings with it added value. Indeed, The

1 The Economist (2017), although we are mindful that this metaphor is not without its problems.

Economist calculates that 1.4 trillion USD of Alphabet (the owner of Google) and Facebook's combined market value of 1.9 trillion USD comes from turning valuable data into even more valuable insight.² Artificial Intelligence (AI) is likely to be a key way in which this added value is created. It is therefore not surprising that nation states which generate lots of data, including the European Union (EU), see the refining process as strategically valuable.

The strategic importance of data, however, stretches further than oil: it is not just a natural resource that can be mined and refined into something valuable.³ Instead, data comprises information that is itself a resource. This information may include (potentially sensitive) content about individuals and organisations deserving of consideration, as its use may lead to an obscuring of accountability in non-transparent and automated ways or, at worst, to unintended harmful and biased effects. For reasons such as these, the twin considerations of value capture and ethics underpin the EU's policy approach to AI, which has come to emphasise ethical considerations, human centricity and trustworthiness both as core values and as strategic imperatives.

In this volume, we zoom in on how the EU's AI policies and guidance have influenced and been adopted by a number of its member states (and Norway, which is part of the European Economic Area). However, EU-level policies are only as influential as the policies they lead to in member states. The way in which member states interpret EU policies, and support national initiatives furthering their goals, is likely to decide whether the EU's strategic focus on human centricity and trustworthiness leads to strategic advantages in addition to ethical approaches.

The importance of trust and trustworthiness were explicitly pointed to in the *European AI strategy*, published in April 2018.⁴ Parallel to the aims

2 The Economist (2020).

3 Arguably, data, in contrast with oil, does not occur naturally but is generated by creating infrastructures that demand interactions so as to generate signals that can be recorded.

4 European Commission (2018a).

of investment in research and innovation and to prepare for socio-economic changes, trust and accountability were specifically addressed under the third pillar ‘ensuring an appropriate ethical and legal framework’. As a result of this strategy, the High-Level Expert Group on Artificial Intelligence (AI HLEG), an independent expert group set up to guide European policy in AI, was set up by the European Commission in June 2018.

Importantly, the European AI Strategy pledged to produce a coordinated plan with member states – and Norway and Switzerland – in order to “maximise the impact of investments at EU and national levels, exchange on the best way for governments to prepare Europeans for the AI transformation and address legal and ethical considerations”.⁵ This *Coordinated Plan*, supporting an AI “made in Europe,” was subsequently published in December 2018, encouraging all member states to develop their national AI strategy by mid-2019, building on the work done at the European level.⁶

Alongside increasing investment, making more data available and fostering talent, the four key areas pointed out in the Coordinated Plan also explicitly included ensuring trust. This was expressed as stressing:

Implementing, on the basis of expert work, clear ethics guidelines for the development and the use of AI in full respect of fundamental rights, with a view to set global ethical standards and be a world leader in ethical, trusted AI.⁷

Even if all member states were not successful in drafting and publishing AI strategies on their own by mid-2019, the Coordinated Plan set a development in motion at national level with regards to these values.⁸

5 European Commission (2018a), p. 19.

6 European Commission (2018b; 2018c).

7 European Commission (2018c), p. 3.

8 For an overview of the AI strategies of the Member States’, see van Roy (2020) and the AI National Strategy Reports prepared by AI Watch in collaboration with the OECD.ai.

In parallel, the AI HLEG⁹ published the influential *Ethics Guidelines for Trustworthy AI* in April 2019, hereinafter the Ethics Guidelines, and the subsequent *Policy and Investment Recommendations for Trustworthy AI* in late June 2019.¹⁰ These two documents, produced by a mix of academic researchers and representatives from both industry and NGOs, were clearly signs of an increased awareness of ethical and value-based concerns surrounding applied AI,¹¹ also indicating a trend on principled ethical and normative statements on AI.¹² The European Commission followed suit with the *White Paper on AI* in February 2020, developing an “approach to excellence and trust”.¹³

The question still remains, however, to what extent this human-centred policy-approach on trustworthy AI at EU level also is reflected in, and influences, member state strategies.

1.1 Purpose: Trustworthy AI as a strategic priority in the Member States?

The key purpose of this report is to analyse to what extent the notions of ethical, human-centred and trustworthy AI clearly proposed at the European level also have influenced the AI strategies at member state level. In order to do so, we focused on a sample, drawing on: Portugal, The Netherlands, Italy, the Czech Republic, Poland, Norway and the Nordics.

The invited analysts have focused primarily on the published documents that have indicated a strategic approach to AI. For some member states this involves several documents, and for some only one. Occasionally, additional information has been gathered through interviews. The timing between EU-level development and the member states’ strategic work on AI is of importance and, which we will see

9 See also the commentary in this volume from one of its members, the AI-researcher and Associate Professor in Computer Science, Fredrik Heintz.

10 AI HLEG (2019a; 2019b).

11 Larsson (2020).

12 Jobin et al. (2019).

13 European Commission (2020).

below, have played out in different ways for the analysed countries. For an overview, see the timeline in Figure 1.

2. EU overarching policies

FIRST OF ALL, the primary EU-level policy documents that have informed the analyses of member state strategies the most are the following three:

- 1 The *Ethics Guidelines for Trustworthy AI*, published by the AI HLEG in April 2019;¹⁴
- 2 The 33 proposed *Policy and Investment Recommendations for Trustworthy AI*, addressed to EU institutions and Member States, published by AI HLEG in June 2019;¹⁵ and
- 3 The European Commission's *White Paper on Artificial Intelligence*, published in February 2020.¹⁶

These sources are naturally closely linked to other documents, as indicated above. For example, in order to move from the AI domain seen as a research area, towards AI seen as a concept to include in normative statements, the definition of AI became particularly important to clarify.¹⁷ This resulted in the AI HLEG drafting and publishing a report on the AI definition.¹⁸ Also, the national strategies that were published before all of the three mentioned in this list, can obviously not have been influenced by them, but may still be of interest to see what sort of notions have been developed in parallel. Additionally, the White Paper is chronologically late in relation to most member state strategies, but is of course pertinent in itself to the extent that it indicates the European vision of AI as an “approach to excellence and trust”.¹⁹ There are also strategic documents at member state level that have had the

14 AI HLEG (2019a).

15 AI HLEG (2019b).

16 European Commission (2020a; cf. 2020b).

17 For an analysis of this transition of the concept and its normative implications, see Larsson (2020).

18 AI HLEG (2019c).

19 European Commission (2020a).

Member State strategies and EU-level policies over time

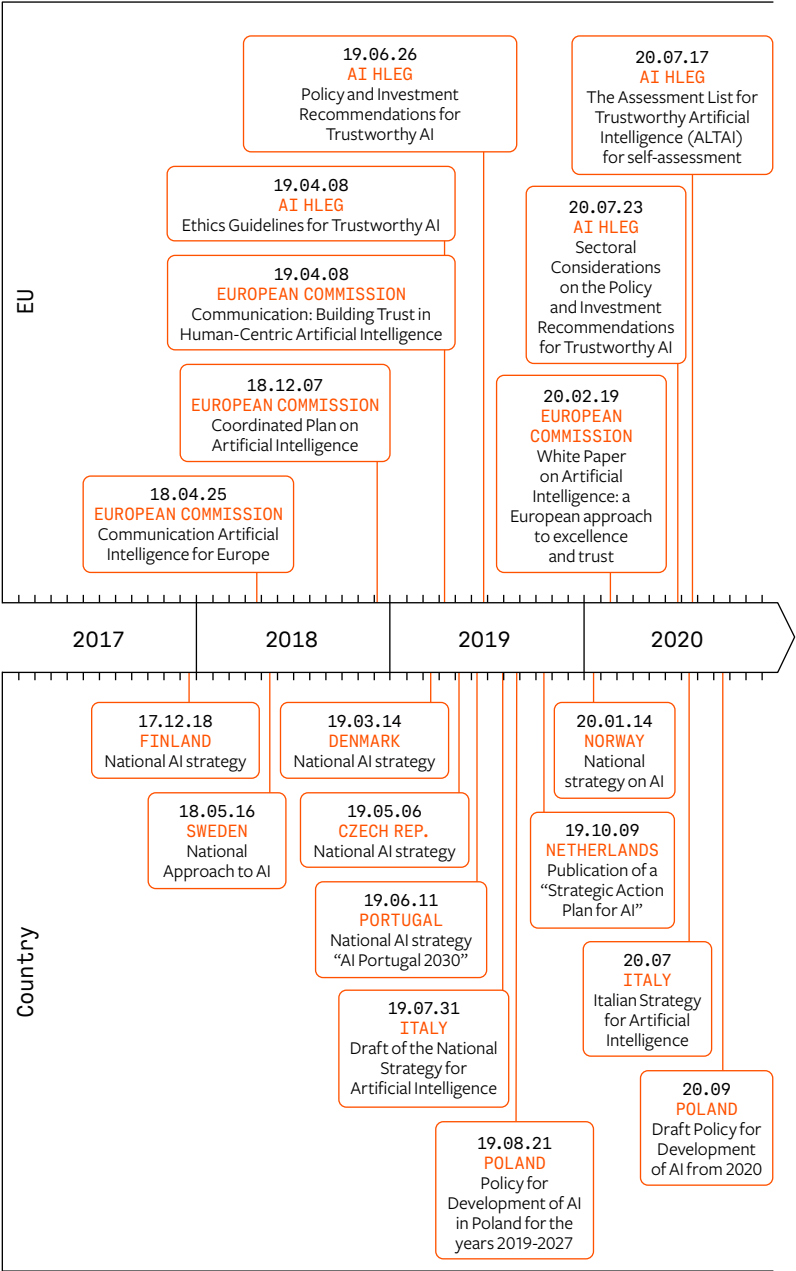


FIGURE 1: Timeline of key EU-level policy documents and sample of Member State strategies.

chance to reflect on the notions of this approach as well as the simultaneously published *European Data Strategy*,²⁰ for example the Polish strategy analysed in chapter 3 below. Furthermore, there are also links that may be of relevance between the persons drafting the EU-level documents and the national discourses on strategies, that may have an impact.

The following presentation is not intended to provide a full description of the key publications, but a sample of the most relevant issues for the purpose of this anthology.

2.1. The Ethics Guidelines

The Ethics Guidelines²¹ are comprised of four levels: (i) a framework stating that trustworthy AI is composed of being lawful, ethical and robust; (ii) ethical foundations for trustworthy AI as found in the respect for human autonomy, prevention of harm, fairness, and explicability; (iii) seven requirements for the realisation of trustworthy AI in deployment, as well as (iv) a “non-exhaustive” assessment list directly organised under these seven requirements. This Assessment list was piloted during the second half of 2019 and was published in July 2020 as the *Assessment List for Trustworthy AI (ALTAI)*, intended for self-evaluation purposes.²²

The seven requirements are arguably what has most clearly influenced strategies on AI, as we shall see below. Those are:

- 1 Human agency and oversight
- 2 Technical robustness and safety
- 3 Privacy and data governance
- 4 Transparency
- 5 Diversity, non-discrimination, and fairness
- 6 Societal and environmental wellbeing
- 7 Accountability

20 European Commission (2020b).

21 A first draft of the Ethics Guidelines was released on 18 December 2018 and was subject to an open consultation which generated feedback from more than 500 contributors. This feedback was used to shape the final version, published on 8 April 2019.

22 AI HLEG (2020a).

As indicated in Fredrik Heintz's commentary, the seven requirements for the realisation of trustworthy AI should be continuously evaluated, and not something to merely be performed once with an expectation that all problems are solved.



FIGURE 2: Seven requirements for the realisation of trustworthy AI.²³

Of interest here is that several of these requirements also map onto well-established legal domains, such as privacy and non-discrimination. As we shall develop below, transparency is a common theme in formalised ethics related to AI, but it is also multifaceted and not necessarily an easily pinpointed terminology.²⁴

²³ AI HLEG (2019a).

²⁴ Larsson & Heintz (2020).

Critique has been voiced towards the Ethics Guidelines in terms of ethical principles lacking the procedural strength of law.²⁵ Relatedly, and in line with the self-regulatory claims of several large digital platforms, concerns have been voiced about allowing representatives of the industry too much control over regulatory issues governing AI.²⁶

2.2. The Policy and Investment Recommendations

The Policy and Investment Recommendations was the AI HLEG's second deliverable and was published on June 26th 2019.²⁷ It comprises 33 points (including several sub-points) divided into eight groups concerning recommendations on human empowerment, the public sector, and research capabilities to data management, educational issues, governance and funding. The document is detailed and provides guidance for many vastly different sectors and topics.

Given that the purpose of this report is to focus on the notions of ethical, human-centred and trustworthy AI, we here address a few sections from the Policy and Investment Recommendations of particular significance. The recommendations take broad understanding of the societal impact of AI, as visible in the section of measuring the societal impact of AI:

- 5.1 Encourage research and development on the impact of AI on individuals and society, including the impact on jobs and work, social systems and structures, equality, democracy, fundamental rights, the rule of law, human intelligence, the development of (cognitive skills of) children.

This can be read as a multidisciplinary call for improving knowledge on the relationship between AI and society. There is a sense of the importance of external scrutiny of AI systems, expressed in the subsequent 5.2. of the recommendations: they advise independent testing of AI

²⁵ Hagendorff, T. (2020); Coeckelberg (2019).

²⁶ For a discussion of this critique and the temporal challenges of the relationship between new technologies and law, see Larsson (2020).

²⁷ AI HLEG (2019b).

systems by civil society organisations and other independent parties. This is in line with calls elsewhere for supervisory methods and competencies among authorities responsible for things like consumer protection.²⁸ Furthermore, aware of contemporary debates around the impact of digital platforms,²⁹ note that aspects of these debates also are reflected in the recommendations. For example:

- 2.2 Commercial surveillance of individuals (particularly consumers) and society should be countered, ensuring that it is strictly in line with fundamental rights such as privacy – also when it concerns “free” services – taking into consideration the effects of alternative business models.

This includes “power asymmetries”³⁰ and is of clear relevance for ongoing revisions in the European competition field. Regulatory recommendations follow further below in the document. Specifically, for consumer protection, for example:

- 27.4 For consumer protection rules: consider the extent to which existing laws have the capacity to safeguard against illegal, unfair, deceptive, exploitative and manipulative practices made possible by AI applications (for instance in the context of chat-bots, include misleading individuals on the objective, purpose and capacity of an AI system) and whether a mandatory consumer protection impact assessment is necessary or desirable.

As shown below in section 4, these points tie into critical perspectives found and developed in research on applied AI and machine learning. The AI HLEG does however not recommend increased regulatory and enforcement capacity,³¹ which is a clear part of interest in the subsequent White Paper on AI, from the European Commission.

28 Larsson (2018).

29 Andersson Schwarz (2017); Larsson & Andersson Schwarz (2018).

30 AI HLEG (2019b), point 2.3.

31 For a critical analysis, see Veale (2020).

2.3. The White Paper on AI

Given that most member state strategies were published before the EU Commission's White Paper on Artificial Intelligence³², they have not been influenced by it. The White Paper is however of relevance here in terms of how it interplays with the work of the AI HLEG and how it points to future regulatory developments, for example linked to notions of risks with AI.

When the White Paper was published on February 19th 2020, it was accompanied by a report on the safety and liability implications of AI, IoT and robotics (as well as a European data strategy,³³ which we return to in section 4 below).³⁴ As stated in the White Paper and elsewhere,³⁵ many of the issues that the trustworthy approach on AI entails are already regulated, for example in data protection and antidiscrimination. The report on safety and liability discusses implications of autonomy and self-learning features of AI-products, particularly with regards to risk assessment.³⁶ This is obviously of relevance for the notion of human-centred AI, that includes human control. Furthermore, the “opacity” and “black box-effect” that some AI-systems may have on the decision-making process is pointed to as an enforcement and accountability problem.

The White Paper consists of two main blocks based on the notion of “ecosystems”; one on excellence and one on trust. This means that there is something of a two-pronged approach: examining the possibilities on the one hand – linked to calls for research, member state collaboration, innovation and increased investments – and the risks or challenges on the other – to ensure trustworthiness, liability, and safety. The latter is of particular interest for this volume's focus on human-centred and trustworthy AI, and is also the one that most clearly relates to regulatory questions of AI.

³² European Commission (2020a).

³³ European Commission (2020b).

³⁴ European Commission (2020e).

³⁵ Larsson (2020).

³⁶ European Commission (2020e), pp. 6-7.

The Commission states that the legislative framework could be improved to address

- ▶ Effective application and enforcement of existing EU and national regulation. This is to say that existing law, in many cases, is fit for purpose but is challenged from the perspective of implementation. The Commission specifically points to a lack of transparency that makes it difficult to identify and prove possible breaches.³⁷
- ▶ The limitation of scope of safety legislation that applies to products and not to services, and therefore in principle not to services based on AI technology.
- ▶ The changing functionality of AI systems, for example for products that rely on frequent software updates of machine learning.
- ▶ The allocation of responsibilities at different places in a supply chain.
- ▶ Changes to the concept of safety, related to for example cybersecurity.

While it is not clear how these insights will be accommodated, the White Paper was opened for a public consultation process that ended 14th of June, 2020, receiving over 1,200 contributions.³⁸

Of particular relevance, and also the object for debate, is the proposed definition of risk in AI, since it is used to indicate the needs of future regulations. Guided by the principle of that “the new regulatory framework for AI should be effective to achieve its objectives while not being excessively prescriptive so that it could create a disproportionate burden”,³⁹ the Commission suggests that high-risk applications are distinguished from all other applications; especially pointing to healthcare, transport, energy and parts of the public sector as sectors where, given the characteristics of the activities typically undertaken, significant risks can be expected to occur. Furthermore, and cumulatively, the AI application would need to have been used in such a way

37 European Commission (2020a) p. 14. For a study on the need for supervisory authorities to improve supervisory methodologies and “algorithmic governance”, see Larsson (2018).

38 European Commission (2020d).

39 European Commission (2020a), p. 17.

that significant risks were likely to arise. The high-risk sector-requirement have received critique,⁴⁰ for example in relation to some of the issues described under the subsection on AI and ethics below (4.1.), as well as the “commercial surveillance” pointed to by the AI HLEG in the Policy and Investment Recommendations.⁴¹ These types of risks are not necessarily found in high-risk sectors. The German government, for example, has called for the proposed risk-classification system in the White Paper to be revised.⁴² This call was likely informed by the more levelled approach on AI risks proposed by the German Data Ethics Commission⁴³ a few months prior to the publication of the White Paper.

3. Contributions of each chapter, by country

IT IS CLEAR that the EU-level policies have had an impact on several of the national strategies. While some countries have explicitly incorporated aspects from the Ethics Guidelines, such as Norway and Portugal, others are predisposed to including questions of trust and transparency, as in the Nordics, or ethics, such as Poland. It is however also clear that the EU’s Ethics Guidelines have had more impact than its Policy and Investment Recommendations. The main results from the analyses are collected here.

3.1. Portugal⁴⁴

In Portugal, policy discourse around AI seems very typically European, in its devotion to human-centred values such as privacy protection, safety, transparency, fairness, and trans-European inclusion. Nevertheless, Pedro Rubim Borges Fortes argues that official Portuguese AI policy is characterised by being quite laconic in its definition of AI, compared to the top-level descriptions of AI provided by

⁴⁰ Dignum et al. (2020).

⁴¹ AI HLEG (2019b), point 2.2.

⁴² Die Bundesregierung (2020).

⁴³ The German Data Ethics Commission (2019).

⁴⁴ Chapter 2, by Pedro Rubim Borges Fortes, Visiting Professor at the Doctoral Programme at the National Law School of the Federal University of Rio de Janeiro and Public Prosecutor at the Attorney General’s Office of Rio de Janeiro.

the AI HLEG. The Portuguese strategy primarily focuses on big data processing and emphasises the importance of innovation, but remains rather silent about the potential role of law, regulation, and consumer protection – especially in comparison with the more general European framework.

Arguably, Portugal could be said to be one of the European countries where technology diffusion and literacy has been at an average level (in this way comparative also to the examples, in this report, of Italy, Poland, and the Czech Republic). The topic of modernisation of public administration is given quite considerable prominence in Portugal's national AI strategy, with an eye on transparency, auditability, privacy protection, and fairness. Likewise, education and civic empowerment is emphasised, focusing especially on the young. It seems clear, from the overview of Portuguese AI policy in this chapter, that the country should neither be seen as being at the cutting edge of AI innovation, nor be seen as a laggard; Portugal presents ambitious plans at being at the forefront of the development of digital skills, and has an impressive track record in terms of conditions for tech development, especially as Lisbon has been the host to the international Web Summit, one of the largest tech events in the world, for the last five years (taking over the role after Dublin passed on it in 2015).

3.2. Poland⁴⁵

With a current majority government that explicitly opposes liberal democracy – despite an economy that has been booming for several years, a population that is increasingly digitally skilled, and with large contingents of the population holding progressive, anti-authoritarian values – Poland finds itself as one of the countries caught in a political-economic paradox. Arguably, ethics is at the core of this Polish national conjuncture, as the country is currently ruled by a very conservative party, founding much of its politics on ostensibly ethical concerns – albeit on authoritarian, populist, and religiously dogmatic

45 Chapter 3, by Kasia Söderlund, LLM, PhD student in Technology and Society, LTH, Lund University, Sweden.

grounds. However, the country's AI policy documents show little trace of overt moralism or authoritarianism; more specifically, a close reading of the documents reveals a dedication to ethical deliberation that hints more towards the principled, calm and measured reasoning of public servants rather than that of religious dogmatic authoritarians. The Polish AI policy embodies classically European enlightenment values, such as trustworthiness and equal rights.

The chapter on the Polish AI strategy therefore argues that ethical dimensions of AI have been taken seriously by Polish policy makers throughout the legislative process. Here it is, in particular, the right to human dignity that is recognised as foundational to any policymaking about the role of technocratic management. The Polish policymakers seem to have taken the Commission's recommendations seriously, and have incorporated the ideas of trustworthy AI at the core of their drafts, which includes rather specific sets of proposed actions and objectives in areas such as: infrastructures for science and expertise, educational initiatives, international cooperation, public-sector procurement for AI solutions and improved coordination of AI development, and digital competence as a way of improving demographic attractiveness to AI-related investment.

3.3. Norway⁴⁶

Demographically speaking, Norway is one of the richest and technologically most advanced countries in Europe, yet its core industrial strength lies primarily in other sectors than software services and management. While information and communications technologies (ICT) is a growing sector in the country (telecoms in particular), ICT remains a rather modest export industry compared to maritime, offshore oil and gas, seafood, mining, and manufacturing, industrial production/manufacturing, and services such as finance/trade and travel/tourism.

In terms of administration and governance, compared to other EEA

46 Chapter 4, by Frans af Malmborg, PhD student in Public Administration at the Department of Political Science and Management at the University of Agder, Norway.

and EU countries, Norway has modelled its AI strategy in line with the European Commission's guidelines. Its strategy emerges against the backdrop of its considerable oil wealth and reputation for already having made considerable progress towards digitalisation seen, for instance, in the European Innovation Scoreboard 2020 ranking Norwegian firms the third most digitised in the EU.

Norway's AI strategy therefore seems to put more focus on the applications and uses of AI rather than the provision of platforms, data centres, and software infrastructures. According to Frans af Malmberg, the chief areas of focus in the Norwegian AI strategy lie in enabling innovation, enabling skills development, and supporting research and development. However, ethics and trustworthiness are seen as an important part of these areas of focus, both because of Norway's own commitment to human rights and human development, but also as a competitive advantage. To quote their former Minister of Digitalisation, Nikolai Astrup: "human-friendly and trustworthy artificial intelligence may prove a vital competitive advantage in today's global competition", indicating a clear influence from the EU-level publications, particularly the AI HLEG.

3.4. The Nordics (Denmark, Finland, Norway, and Sweden)⁴⁷

The chapter focuses on and compares the official AI policies of four Nordic countries: Denmark, Finland, Norway, and Sweden. These countries share common traits of relevance for the European AI policies, such as their (arguably) relatively similar cultures, with high levels of openness in society, transparency with regards to information access, and trust. By examining four national strategic plans for developing AI, Cory Robinson explores the distinctive differences in how Nordic countries position themselves using their unique cultural values as business principles to support development and deploying AI technologies throughout their societies.

47 Chapter 5, by Cory Robinson, PhD and Senior Lecturer/Assistant Professor of Communication Design at Linköping University in Sweden.

While their national policies differ in clarity and specificity (where Sweden stands out as the least ambitious), all of these countries embody three core values: openness, transparency, and trust. Being rich and already technologically adept countries with strong engineering cultures, the Nordic countries seem to agree on how they, while not being the world leaders in *building* AI technologies, might prove to be vanguard nations in terms of *applying* AI, both in their public administrations and in the private sector. Each country seems to recognise, in their national policies, how knowledge about one's strategic weakness can become an advantage when mapping out those potential uses of AI that might be the most beneficial to both society in general and its individual citizens. The fact that many of these Nordic countries already seem to have had cases of potential overreach by public authorities applying datafied or even AI-based solutions to local bureaucracy indicates that AI is already a critical issue to be dealt with, not merely a future prospect.

The analysis of the Nordic policy documents finds that while openness is a cultural value that is lauded in these countries, it is largely absent in the policy documents, while the values of transparency and trust manifest explicitly or through related concepts. Robinson's chapter thus confirms several of Malmberg's initial observations about Norway in specific. Each Nordic country presents vastly different policy documents, and particularly the government policies of Denmark, Finland and Norway should, according to Robinson, be commended for tackling complex technical issues and informing how AI will potentially impact society – including how citizens will interact with government, and receive services from those very governments.

3.5. The Netherlands⁴⁸

On the face of it, the Netherlands is well-positioned to take advantage of AI advances in Europe. It is a nation that makes eager use of public sector data in fields like social security and that ranks highly on measures of digitalisation and research. The country has a three-pronged

48 Chapter 6, by Katja de Vries, an Assistant Professor in Public Law at Uppsala University in Sweden.

strategy, in which it plans to capitalise on existing economic and social strengths, create the right conditions for AI, and support the foundations of AI research. Trust is particularly visible in the third of these prongs; not only is it explicitly mentioned in the country's AI strategy, it is implicit in their stated commitments to improving data security and promoting human rights and values.

According to Katja de Vries, however, its implementation of AI policy, though ambitious, is not without its problems. The first of these lies in the fact that AI policies have been promulgated by three different state agencies – leading to concerns around overlap, fragmentation and a lack of coherent strategic vision. The second of these lies in the Netherlands' recent history of failing to be fully transparent when using algorithms in the public services, in two very public cases of inaccurate algorithmic tools used to detect benefits fraud. Lastly, the funding for the Netherlands' ambitious AI agenda is largely made up of a combination of redeployed existing commitments, and the hope that significant sums can be drawn from the private sector. This calls into question the financial sustainability of the proposed initiatives.

3.6. Italy⁴⁹

Like Poland, Italy is a considerably large European country with a strong heritage in public administration. Its AI strategy indicates that many of its ministerial departments are staffed by civil servants who employ a principled style of policymaking that is already historically established in distinctly European understandings of civic values and human rights. According to Francesco Cappelletti, the Italian strategy emphasises several things: Development of core infrastructural components for digitalisation, in order to serve both public administration and industrial interests. Research and partnerships between institutional actors is emphasised, so as to not only support the country's many small and medium-sized industries, but also bolster larger industrial actors, and to reform public administration. An emphasis is

49 Chapter 7, by Francesco Cappelletti, Research Fellow at European Liberal Forum; member of Fondazione Luigi Einaudi in Italy.

put on creating an integrated ecosystem that manages to serve the very heterogeneous domestic geography of the country (with vast demographic and industrial differences between south and north and different regions). Like in many of the other national strategies, much weight is also put on the advancement of skills required to build new digital infrastructure; both civilian literacy and industrial know-how. The question marks that are raised by Cappelletti have mainly to do with the actual implementation of these grand plans; Italy is well known for its political tumultuousness and its often very slow legislative processes.

3.7. Czech Republic⁵⁰

The overarching sentiment of the national AI strategy of the Czech Republic is the devotion to national economic growth and competitiveness, as this underpins most of the constituent elements of its national AI policy. The humanistic values espoused by the European Commission are in this way addressed more indirectly than directly, as the main strands of the national policy all converge back to core economic concerns (e.g. industrial research and development, financing, and human capital) while more human-oriented aims (e.g. education, regulation, and international cooperation) are understood also through this economic lens.

Seen as a whole, the Czech strategy ought to be applauded for its overarching multi-stakeholder design. That said, since the AI strategy of the Czech Republic is coordinated by the Ministry of Industry and Trade, with interim progress reports submitted yearly to the Steering Committee and to the government, it is instructive to understand its NAIS primarily through the lens of industrial and trade policy.

This focus on business-oriented discourse is not surprising, given that the Czech Republic is a rather small central-European country, in the midst of a postindustrial transition where the comparatively low-wage country excels, particularly, in technology-intensive manufacturing and engineering, where digital services (e.g. software management,

50 Chapter 8, by Anas Zaza, independent researcher and External Analyst with the Institute of Politics and Society in the Czech Republic.

back-end operations) seem to make up an increasing part of its productive capacity. The AI policy explicitly speaks of “Centres of Excellence in AI research,” “test centres,” and “digital innovation hubs,” for example.

4. Parallel developments, and looking to the future

THESE AI POLICIES have, of course, not emerged from nowhere. There have been a considerable number of parallel developments that intersect with AI policy – chiefly when it comes to data. Here, we discuss (1) parallel AI ethics discussions, (2) the socio-political context of AI policy emergence in the EU, as well as related data policies that have implications for AI, specifically in respect of (3) the public sector, and (4) the need for further regulations.

4.1. AI and ethics

The European approach on trustworthy AI has evidently been developed in close tandem with an increased academic and corporate awareness of not only the possibilities inherent in AI and machine learning but also their perils.⁵¹ Via debates spurred by illustrative cases on racial bias in risk assessment for recidivism,⁵² and studies revealing commercial uses of facial recognition systems are less precise when it comes to women, dark skin,⁵³ and for pedestrian detection,⁵⁴ new research fields have emerged focusing on the intersection of fairness and accountability for autonomous systems.⁵⁵ Research is also focusing on issues in the intersection between society and AI,⁵⁶ including responsible AI,⁵⁷ its link to ethics,⁵⁸ and issues of regulation⁵⁹ and governance.⁶⁰

51 For inventories, see Larsson et al. (2019), Whittaker et al. (2018).

52 ProPublica (2016); Eubanks (2017): cf. Fortes (2020).

53 Buolamwini & Gebbru (2018).

54 Wilson, Hoffman & Morgenstern (2019).

55 Cf., the traditional AI conferences starting to include ethics and society, such as the AAAI/ACM conference on Artificial Intelligence and Ethics and Society, AIES, or entirely new ventures like the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT).

56 For example, the 10-year Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society.

57 Dignum (2019).

58 Dubber, Pasquale & Das (2020).

59 Yeung & Lodge (2019).

60 Cf. Ala-Pietilä & Smuha (2020); Fjeld et al. (2019); Larsson (2020).

The governance issues of AI have spurred many guidelines and principled stances on AI from both governments, NGOs, as well as companies.⁶¹ One study of the global AI ethics landscape identified 84 documents containing ethical principles or guidelines for AI.⁶² They were collected up until April 2019, with 88 percent released after 2016. Interestingly enough, out of the 84 documents, they found a global convergence around the five ethical principles of (1) transparency, (2) justice and fairness, (3) non-maleficence, (4) responsibility, and (5) privacy. However, they also noted that there seem to be considerable differences in how these principles are interpreted, why they are considered important, what issue, domain or actors they relate to, and how they should be implemented.⁶³

In another recent analysis of AI ethics policies around the world,⁶⁴ only two of the 22 analysed ethics policies explicitly addressed aspects like “democratic control, governance and political deliberation of AI systems”⁶⁵ and explicitly prohibited normative stipulations by AI systems of particular lifestyle choices.⁶⁶ This leads to an interesting observation in the study:

Hardly any guideline discusses the possibility for political abuse of AI systems in the context of automated propaganda, bots, fake news, deep fakes, micro targeting, election fraud, and the like. What [was] also largely absent from most guidelines [was] the issue of a lack in diversity within the AI community.⁶⁷

Furthermore, none of the 22 surveyed guidelines detailed whether systems of algorithmic decision-making “are superior or inferior,

61 Fjeld et al. (2020).

62 Jobin, Ienca & Vayena (2019).

63 Ibid.

64 Hagendorff (2020).

65 Ibid., p. 105.

66 This latter aspect is seen as particularly interesting, given the relatively large global interest in a proposed Chinese credit scoring system which is said to have such prescriptive features; see Engelmann et al. (2019).

67 Hagendorff (2020), p. 105.

respectively, to human decision routines [and] virtually no guideline [dealt] with the ‘hidden’ social and ecological costs of AI systems”.⁶⁸ What did appear in most policies, however, according to the same analysis, are aspects of *fairness*, *accountability*, *transparency*, and *privacy*. Such values appeared in around 80 percent of the guidelines surveyed, providing what the author sees as a set of “minimal requirements for building and using an ‘ethically sound’ AI system”.⁶⁹ What is striking, argued for in this study, is that these more frequently mentioned aspects are such that they also tend to have already existing technical fixes, or they are understood to be relatively easily addressable by way of technical solutions.

Consequently, *transparency* is one of the most prevalent principles and challenges linked to AI deployment in the current literature. In the larger study on guidelines, it was featured in 73 out of the 84 sources.⁷⁰ It is part of the ethical foundation put forward by the AI HLEG, in terms of *explicability*, and one of the seven requirements for the realisation of trustworthy AI. Furthermore, *AI transparency* has been shown to be a particularly multifaceted concept requiring balancing of interests on top of the challenges with “black box” algorithms and opaque AI-systems.⁷¹ The European Commission also states in the White Paper that transparency, traceability, and human oversight are not specifically covered under current legislation in several economic sectors.⁷² The lack of transparency, the Commission brings forward, makes it “difficult to identify and prove possible breaches of laws, including legal provisions that protect fundamental rights, attribute liability and meet the conditions to claim compensation.”⁷³

4.2. Trustworthy AI as a competitive advantage

Researchers and experts increasingly point to China’s dominance in

68 Ibid., p. 105.

69 Ibid., p. 103.

70 Jobin, Ienca & Vayena (2019), p. 391.

71 Larsson & Heintz (2020).

72 European Commission (2020a).

73 Ibid., p. 14.

all things data – and especially when it comes to AI.⁷⁴ Although China boasts a world-leading firm, Alibaba, many of the other world-leading AI firms hail from the United States, including: Amazon Web Services, Google Cloud, and Microsoft Azure. The advances that these countries have made in AI could be said not only to set the bar against which the EU must compete, but also pits EU countries against the US and China when it comes to AI. Indeed, commentators have pointed out that none of today’s tech giants hail from the EU – it comes as no surprise that EU countries want to change that when it comes to AI.

Why is it that EU countries want to be AI frontrunners? The first reason is around competitiveness. Within the “data as oil” metaphor, it bears mentioning that refineries that “add value” to a natural resource are not only a source of revenue for states, but also help and encourage an economy to diversify.⁷⁵ Control of both a natural resource and downstream capabilities is good for a country’s international competitiveness, and countries without oil reserves have long been concerned about their dependence on those who do downstream work. Translate this into digital-speak, and we see that digital services, dependent on data, are significant employers, sources of innovation, and sources of national pride. When it comes to the US and China, this competition for supremacy in AI technologies and the benefits that accrue has already been labelled a “Technology War”.⁷⁶

The second reason is related to the flexibility and borderlessness of digital infrastructure, and data in particular. Although many laws exist to protect the integrity of EU residents and citizens from the overreach of firms and governments abroad,⁷⁷ there is little that EU countries can do to stop

74 Johnson (2018).

75 See, e.g., Baur (2014).

76 Ernst (2020).

77 For instance, rules like the General Data Protection Regulation (2016), better known as simply “GDPR”.

it in practice.⁷⁸ Data can easily be moved across borders with little to no resistance, and that innovation in AI and data insight typically requires little more than a smart individual and a laptop. This means that the best strategy for protecting individuals' integrity is not just to introduce punishments (through legislation) when unethical practices occur, but also to offer a carrot to encourage innovative firms to set up shop within the EU, making them more likely to follow nation states' rules and regulations.

Lastly, it could be argued that policies that promote trustworthiness may even improve the quality of the underlying data, and thus innovation and analyses. Policy initiatives around data, including ones pertaining to AI, rest on techno-optimist assumptions that can be summed up with quips such as "the data never lies".⁷⁹ Whether or not this is true, large data sets are notoriously full of gaps that make them potentially unreliable. Data collected online, for instance, may be biased precisely because users know that they are being watched – and adjust their behaviour accordingly. Data online is also often a by-product of other activities, meaning that it is likely not comprehensive, leading to gaps and errors.⁸⁰ Making analysis processes more transparent and trustworthy is likely to engender trust in the data collection process and analysis process. This may mean that individuals are more likely to allow their data to be collected and processed by firms, leading to fewer gaps and thus fewer errors. Indeed, new initiatives to promote EU markets for private and non-private data highlight this strategy,⁸¹ and are the starting points for business models that price data more accurately – and allow Europeans to have a share in the windfall.

78 At the time of writing, when it comes to legal tests of the GDPR there is considerable development, e.g. the work by Austrian lawyer and activist Max Schrems, who is contesting the legality of transatlantic data transfers. There are several ongoing legal cases that might provide a legal basis for forcing both US and European businesses to halt or considerably change their ways of making business in the EU. Conversely, these developments might also lead to an overhaul of the European laws or a re-appropriation of the ways they are interpreted and implemented.

79 e.g. boyd and Crawford (2012).

80 Such as with western bias of the ImageNet image database, leading to less precision for cultural attributes in the east, such as Indian wedding dresses that don't resemble the American ones, see Shankar et al. (2017), discussed in Larsson (2019).

81 European Commission (2020b).

The EU achieves all three goals simultaneously: policies emphasising trustworthiness are not only likely to pre-empt possible conflicts around data being exploited, which would threaten AI research and commercialisation, but allow nation states to protect sensitive data.

4.3. AI in the Public Sector

The EU strategies for AI and data developments do not only extend to research and private sector innovation. Instead, parallel EU developments also point to a movement towards support and encouragement for AI use in the public sector. This is particularly visible in the sectoral considerations that the AI HLEG published in 2020, including questions of public procurement, public responsibility for data and workers affected by AI, and automated decision making in areas like law enforcement and healthcare. Almost half (15 out of 33) of the recommendations put forward by the AI HLEG in June 2019 were dedicated to public services,⁸² including the emphasis on the strategic use of public procurement to promote trustworthy AI; and the need to safeguard fundamental rights in delivering AI-based public services.

The public sector can make strategic use of public procurement to foster responsible innovation, as well as steering it towards tackling societal challenges and the development of trustworthy AI solutions. Moreover, by delivering higher quality, more targeted and thus more effective services to individuals and groups where appropriate, it can also act as a catalyst for innovation and growth.⁸³

So, public procurement, according to the AI HLEG, holds some of the more important keys for ensuring trustworthy and human-centred AI in the public sector. The expert group also recommends that public authorities play an active role in not just promoting and supporting AI developments, but also supporting those adversely affected by AI, ensuring that workers are included through consultations around AI implementations and where AI makes automated decisions. Crucially,

⁸² AI HLEG (2019b).

⁸³ AI HLEG (2019b), p. 9.

the AI HLEG⁸⁴ has sections devoted to both making sure that the benefits of AI accrue to all (Section 4), and that the public sector should measure and monitor the effects of AI on society (Section 5).

Linked to this imperative to measure and monitor the societal effects of AI, is the recommendation that public actors should get involved in areas around AI bias and undue prejudice (Section 12), which have already been seen in applications of AI in the private sector, and as are mentioned in this anthology's chapter 6, on AI policy in the Netherlands.

The subsequent sectoral considerations, published in July 2020, is an outcome of the Policy and Investment Recommendations being discussed in workshops with invited experts and stakeholders. The workshops led to a number of additional observations and proposals, argued by the AI HLEG to warrant particular attention from EU policymakers in the time to come. Here we account for a sample of these additional observations that we find are of particular relevance for the utilisation of AI in the public and/or health sector. For example, they put additional emphasis on the importance of transparency in the sense that "AI-enabled e-Government services should be accompanied by adequate arrangements in terms of accountability and traceability, enabling ex post verification".⁸⁵ Furthermore, the Sectoral Considerations advocate:

- ▶ Promote data and algorithmic literacy amongst the public administration;
- ▶ Invest in upskilling and reskilling to harness the potential of Trustworthy AI in healthcare;
- ▶ Availability of high quality health data is key for building Trustworthy AI solutions in Europe that can improve public health outputs and help manage patients efficiently;
- ▶ There are concerns on a possible trade-off between access to quality treatment and privacy, as well as possible discrimination as a result of AI use: AI use in the healthcare sector has to be thoroughly and independently monitored.

⁸⁴ AI HLEG (2019b).

⁸⁵ AI HLEG (2020b), p. 9.

Lastly, regarding deployment of AI-systems for the specific law enforcement and justice contexts, the AI HLEG presents a somewhat cautious approach, arguing that “deployment at greater scale generates risks and opportunities that are not yet fully understood”.⁸⁶ The expert group therefore calls for a “wide-spread policy debate in Europe (and beyond) on the development, use and impact of AI-assisted and AI-enabled decision-making systems in justice and law enforcement.”

4.4. Regulatory development

Several of the seven points in the Ethics Guidelines, indicated as necessary for a realisation of trustworthy AI, are already highly regulated, such as anti-discrimination and data protection.⁸⁷ Nevertheless, it is likely that further regulations will be needed to support AI research, development and commercialisation. These development lines may be viewed in a more direct AI-related sense, as well as in a more indirect sense. In the more direct sense, the White Paper, outlined above, has pointed to enforcement challenges, issues of product liability for adaptive and autonomous technologies, as well as the distribution of accountability in complex settings.⁸⁸ The enforcement issues clearly relate to transparency in terms of traceability and explainability of AI-systems remarked on by the AI HLEG.⁸⁹

In the broader, more indirect sense, regulatory issues on the horizon relate to parallel processes of competition regulation for large digital platforms,⁹⁰ such as the Digital Services Act and ongoing cases against alleged misuses of dominant position. Much of the scalable management of the platforms has come to rely on machine learning and versions of AI, oftentimes hidden for scrutiny behind proprietary setups.⁹¹ Some of attached challenges relate to how automated content moderation plays out, or how complex and automated ad-market

⁸⁶ Ibid., p. 11.

⁸⁷ Larsson (2020).

⁸⁸ European Commission (2020a).

⁸⁹ AI HLEG (2019a; 2019b; 2020a).

⁹⁰ Crémer et al. (2019).

⁹¹ Cf. Andersson Schwarz (2017); Pasquale (2015).

infrastructures overrides data protection or risks bringing consumer manipulation to the market.⁹²

Furthermore, and as with most things technology-focused, ongoing policy discussions around AI will struggle to keep up as technological advances proceed apace.⁹³ It is therefore significant that the EU has tried to keep pace with a number of likely consequences of data and AI. These parallel regulatory frameworks are designed to bolster the position of the EU as a human-centric and trustworthy bastion of AI.

The first of these, the Open Data Directive,⁹⁴ seeks to make as much public data as possible available for analysis and innovation. The public data envisioned here is that which is held by the public sector bodies in the member states, such as: ministries, state agencies and municipalities, as well as organisations funded mostly by or under the control of public authorities.

The second of these is the appointment of a High-Level Expert Group on Business to Government data sharing (B2G HLEG). In their final report,⁹⁵ the B2G HLEG emphasised the potential that private-public partnerships around data could have for innovation. At the same time, they highlighted how structures around governance and security were at the centre of making such initiatives work.

Finally, the EU has released a considered overarching European Data strategy,⁹⁶ in which it envisions the EU as an environment in which an abundance of data is available, both from public entities and through envisioned markets for private data. This strategy not only emphasises the importance of trustworthiness and security, but also the flow of data between EU countries, with harmonised interfaces and regulations key to making this happen.

92 Larsson & Andersson Schwarz (2018).

93 Cf. Larsson (2020).

94 Directive (EU) 2019/1024.

95 European Commission (2020c).

96 European Commission (2020b)

5. Conclusions

EUROPEAN AI POLICY development takes place in a complex landscape. It is clear that the technical field of industrial development in machine learning and AI is dominated by two superpowers, namely the USA and China, and that European policymaking has to take relationships with these power blocs into account. Notably, AI development is also dominated by very large, transnational corporations whose operations and reach are on a scale that is on par with the largest companies in the history of the world, like the Saudi Aramco oil company in our time and the Dutch East India Company in its time. It is not surprising that many of these companies are so large as to be seen as akin to national jurisdictions.⁹⁷ Moreover, since the EU is a confederation of sovereign member states, each national policy would also have to take into account the relations to other European nations, as well as the jointly coordinated plans outlined on the EU Commission level. While EU member states of course share many democratic values and administrative traditions, there are still significant differences between different EU member states, both in terms of administrative traditions and in development levels – which affects states’ ability to implement and benefit from AI.

As we have seen, one of the key priorities of the European Commission’s Coordinated Plan on AI was to encourage member states to develop their own national AI strategies by the end of 2019. In our collaborative overview and analysis of a number of these national strategies, we find clear evidence that the Ethics Guidelines have made a substantial impact on the strategies released after their publication. The Policy and Investment Recommendations, however, seem to have had less impact. The White Paper is corresponding to some of the content found in the national member states, but was in general published after the drafting of most strategies.

97 Webb (2019). See also Desjardins (2017). Legal scholar Frank Pasquale (2018) has likened the jurisdictions enforced inside the systems operated by corporations such as Apple, Google, or Amazon to a form of “functional sovereignty.”

Are we to see these countries as representative, what can we glean from our general endeavour of mapping these countries? Are some of them outliers? While our ambition has not been to be exhaustive in terms of being able to fully account for the EU-level impact in all member states, we do see analytical strands of relevance for the larger governance issue in the relationship between the EU-level and member states.

FIRSTLY, state policies start from different points of development. The Italian strategy, for example, observes that both its population and its industrial landscape need to be updated in accordance with the possibilities and challenges afforded by AI technology; while the Polish strategy rather modestly suggests that its population is more digitally competent than might be expected, but that steps could be taken towards an even more digitally adept population. The Czech strategy emphasises the need to cater for industrial development and the institutional and formal needs put forward by business interests, while the Portuguese strategy emphasises obtaining the necessary knowledge and skills to prosper in an AI-enabled world.

In contrast, many of the Nordic countries have already embraced datafication and implemented it both in their public administration (with national solutions for electronic identification and payments having been commonplace in many Nordic countries for years) and among private businesses (for instance, Klarna is an expansive fintech actor from Sweden that purports to use AI technologies in its infrastructural back-end). A less flattering implication of this might be the fact that in these Nordic countries (and in the Netherlands), there have already been several examples of potential overreach, as is noted in chapter 5, where AI technologies have been hurriedly applied to public-sector decision making in sensitive areas, e.g. family welfare and social benefits.

SECONDLY, given the stated objective that the EU should be able to compete with the US and China, and that data (and analysis) should easily flow between EU member states, it is critical that there is

harmonisation of national legislation. In some ways, the Norwegian documentation appears to be more discursively aligned to the EU Commission policy than some of the other Nordic countries, despite Norway not being an EU member state. This might be explained by the timing of the Norwegian documentation: when the Norwegian policy was formulated, the AI HLEG policy had already been published, while this was not the case for the other Nordic countries.

We see that AI policy in the Nordic countries can be summarised as being value-laden but where actual implementations of AI technology (which are sometimes already rather far ahead, in this region) at the same time lays bare a pragmatism which puts the original, abstract declarations of “trust,” “openness,” and “transparency” to the test. Here, one could either regard empirical reality as a counterpoint where such values are sometimes not fulfilled – or, one could make the more optimistic reading that public debates where actual implementations are being questioned are actually good signs that these very values are being pursued.

IN CONCLUSION, it is clear that the member states need to be the engines driving the pragmatic changes in the fields of trustworthy and human-centred AI as applied in society, on markets, before citizens and consumers. It is natural that the different national AI strategies unpacked in this anthology all partially mirror cultures of innovation and technology management in each country. However, this is both a strength and a weakness: while national interpretations of EU directives do much to make them more relevant for local conditions, the risk is that states’ strategies begin to diverge as a result, making the overall effect one of fragmentation.

References

- AAAI / ACM conference on Artificial Intelligence, Ethics and Society. (2020) <https://www.aies-conference.com/2021/>
- ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT). (2020) <https://faccconference.org>
- AI HLEG (2019a). Ethics Guidelines for Trustworthy Artificial Intelligence, Brussels: European Commission.
- AI HLEG (2019b). Policy and Investment Recommendations for Trustworthy Artificial Intelligence, Brussels: European Commission.
- AI HLEG (2019c). A Definition of AI: Main Capabilities and Disciplines: Definition Developed for the Purpose of the AI HLEG's Deliverables, Brussels: European Commission.
- AI HLEG (2020a). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment.
- AI HLEG (2020b). Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence.
- AI Watch & OECD.ai (ongoing). AI National Strategy Reports: https://ec.europa.eu/knowledge4policy/ai-watch/national-strategies-artificial-intelligence_en#ainationalstrategyreports
- Ala-Pietilä, P & Smuha, N.A (2020). "A Framework for Global Cooperation on Artificial Intelligence and its Governance." Forthcoming in *Reflections of AI for Humanity*, coord. by B. Braunschweig and M. Ghallab, Springer (2021).
- Andersson Schwarz, J. (2017). Platform logic: An interdisciplinary approach to the platform based economy. *Policy & Internet*, 9(4), 374-394.
- Baur, S. (2014). Refining oil—A way out of the resource curse. *London School of Economics International Development Working paper series* No.14-158.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.
- Die Bundesregierung (2020). *Stellungnahme der Bundesregierung der Bundesrepublik Deutschland zum Weißbuch zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen*. 29 June 2020, COM(2020) 65 final.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91).
- Coeckelbergh, M. (2019). Artificial Intelligence: Some Ethical Issues and Regulatory Challenges. *Technology and Regulation* 31-4.
- Crémer, J., de Montjoye, Y., & Schweitzer, H. (2019). *Competition policy for the Digital Era: A report for the European Commission*. European Commission.
- Desjardins, Jeff (2017). "How today's tech giants compare to the massive companies of empires past". Business Insider. <https://www.businessinsider.com/how-todays-tech-giants-compare-to-massive-companies-of-empires-past-2017-12>
- Dignum, V. (2019). *Responsible Artificial Intelligence. How to develop and use AI in a responsible way*. Cham, Switzerland: Springer.
- Dignum, V., Muller, C. & Theodorou, A. (2020). First Analysis of the EU Whitepaper on AI. ALLAI. <https://allai.nl/first-analysis-of-the-eu-whitepaper-on-ai/>
- Directive (EU) 2019/1024 (2019). The European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG

Dubber, M., Pasquale, F., & Das, S. (Eds.) (2020). *The Oxford Handbook of Ethics of AI*. Oxford University Press.

The Economist (2017). Regulating the internet giants: The world's most valuable resource is no longer oil, but data. <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

The Economist (2020). Who owns the web's data? <https://www.economist.com/business/2020/10/22/who-owns-the-webs-data>

Engelmann, S., Chen, M., Fischer, F., Kao, C., Grossklags, J. (2019). Clear sanctions, vague rewards: How China's social credit system currently defines "Good" and "Bad" behavior. In: *Proceedings of the conference on fairness, accountability, and transparency—FAT '19*, pp. 69–78.

Ernst, D. (2020). Competing in Artificial Intelligence Chips: China's Challenge amid Technology War. *Centre for International Governance Innovation, Special Report*. https://www.cigionline.org/sites/default/files/documents/Competing%20in%20Artificial%20Intelligence%20Chips%20-%20Dieter%20Ernst_web.pdf

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York, NY: St Martin's Press.

European Commission (2020a). *White paper. On artificial intelligence — a European approach to excellence and trust*. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission (2020b). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *A European strategy for data*. <https://op.europa.eu/en/publication-detail/-/publication/ac9cd214-53c6-11ea-aece-01aa75ed71a1/language-en>

European Commission (2020c). Towards a European strategy on business-to-government data sharing for the public interest https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64954

European Commission (2020d). White Paper on Artificial Intelligence - a European Approach, consultation. <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12270-White-Paper-on-Artificial-Intelligence-a-European-Approach/public-consultation>

European Commission (2020e). Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. COM/2020/64 final.

European Commission (2018a). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe, COM(2018) 237 final, Brussels: European Commission.

European Commission (2018b). Coordinated Plan on Artificial Intelligence (COM(2018) 795 final).

European Commission (2018c). ANNEX to the Communication from the Commission to the European Parliament, The European Council, The Council, The European Economic and Social Committee and the Committee of the Regions. Coordinated Plan on Artificial Intelligence.

Fjeld, J., Hillgoss, H., Achten, N., Levy Daniel, M., Feldman, J., & Kagay, S. (2019). *Principled Artificial Intelligence*. Berkman Center for Internet & Society, Cambridge.

Fortes, P. R. B. (2020). Paths to Digital Justice: Judicial Robots, Algorithmic Decision-Making, and Due Process. *Asian Journal of Law and Society*, 1-17.

General Data Protection Regulation (2016). 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.

The German Data Ethics Commission (2019). *Opinion of the Data Ethics Commission*. <https://datenethikkommission.de/en/>

Hagendorff, Thilo (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30: 99–120.

Jobin, A., Ienca, M. & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence* 1: 389–99.

Johnson, K. (2018). China could lead world in AI research in coming years, Elsevier report finds. <https://venturebeat.com/2018/12/11/china-could-lead-world-in-ai-research-in-coming-years-elsevier-report-finds/>

Larsson, S. (2018). Algorithmic governance and the need for consumer empowerment in data-driven markets. *Internet Policy Review*, 7(2), 1–13.

Larsson, S. (2019). The socio-legal relevance of artificial intelligence. *Droit et Société* (3): 573–593.

Larsson, S. (2020). On the Governance of Artificial Intelligence through Ethics Guidelines. *Asian. Journal of Law and Society*, 1–23.

Larsson, S., & Andersson Schwarz, J. (Eds.) (2018). *Developing Platform Economies: A European Policy Landscape*. Brussels: European Liberal Forum asbl.

Larsson, S., Anneroth, M., Felländer, A., Felländer-Tsai, L., Heintz, F., & Ångström, R. C. (2019). *Sustainable AI: An inventory of the state of knowledge of ethical, social, and legal challenges related to artificial intelligence*. Stockholm: AI Sustainability Center.

Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, 9(2).

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pasquale, Frank (2018). “Tech Platforms and the Knowledge Problem”. *American Affairs*, 11 (2): 3–16. Summer 2018. <https://americanaffairsjournal.org/2018/05/tech-platforms-and-the-knowledge-problem/>

ProPublica (2016). “Machine bias”, by Angwin, J., Larson, J., Mattu, S., & Kirchner, L. See <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Shankar, S., Halpern, Y., Breck, E., Atwood, J., Wilson, J., & Sculley, D. (2017). No classification without representation: Assessing geodiversity issues in open data sets for the developing world. *arXiv preprint arXiv:1711.08536*.

van Roy, V (2020). AI Watch - National strategies on Artificial Intelligence: A European perspective in 2019. EUR 30102 EN, Publications Office of the European Union, Luxembourg.

Veale, M. (2020). A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence. *European Journal of Risk Regulation* 1–8.

Webb, Amy (2019). *The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity*. New York, NY: PublicAffairs.

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., Myers West, S., Richardson, R., Shultz, J., & Schwartz, O. (2018). *AI now report 2018*. New York: AI Now Institute at New York University.

Wilson, B., Hoffman, J., & Morgenstern, J. (2019). Predictive inequity in object detection. *arXiv preprint arXiv:1902.11097*.

Yeung, K., & Lodge, M. (Eds.). (2019). *Algorithmic regulation*. Oxford University Press.

AI policy in Portugal

Ambitious, yet laconic about legal routes towards trustworthy AI

Pedro Rubim Borges Fortes

Visiting Professor at the Doctoral Programme at the National Law School of the Federal University of Rio de Janeiro and Public Prosecutor at the Attorney General's Office of Rio de Janeiro. DPHIL (Oxford), JSM (Stanford), LLM (Harvard), MBE (Coppe-UFRJ), BA (PUC-Rio), LLB (UFRJ).

Summary

THE PORTUGUESE AI national strategy originated from a governmental initiative established in 2016 and launched on April 3rd 2017 as a national initiative on Digital Skills (INCoDe.2030). After the EU Commission decided to adopt a high-level AI plan in April 2018, the Portuguese government elaborated its national plan according to the EU guidelines. Preparation of the strategy started in October 2018 and the office of the INCoDe.2030 promoted the AI PORTUGAL 2030 report which was released on June 11th 2019. Portuguese authorities hold that AI should strengthen societal robustness and be ethical-by-design, highlighting the importance of Human-Centered European values and the guarantee of privacy protection, safety, transparency, fairness, and inclusion within the European space. The

AI PORTUGAL 2030 report warns that vulnerable individuals may be screened and targeted with false information and that inclusion also implies digital autonomy, safety, and privacy. Likewise, education is emphasised, focusing especially on the young. At the core of the Portuguese national strategy is civic empowerment in terms of necessary knowledge, skills, and means to prosper in an AI-enabled world. The detailed vision of the strategy is influenced by the recommendations from the EU Commission’s High-Level Expert Group on Artificial Intelligence (AI HLEG), but in comparison to these, the AI PORTUGAL 2030 report provides more generic references to transparency without the same level of detail. The national AI strategy also focuses on modernisation of public administration, with an eye on transparency, auditability, privacy protection, and fairness. In contrast to the “thick” definition of AI adopted by the AI HLEG, the Portuguese adopted a “thin” definition as an umbrella concept for big data processing, emphasising the importance of innovation, but remaining laconic about the potential role of law, regulation, and consumer protection. Based on the concrete efforts to develop the Portuguese digital ecosystem, the EU should consider Portugal as a potential host country for the lighthouse centre of research and innovation for AI in Europe, as envisioned by the White Paper.

1. Introduction

THE PRESENT CHAPTER consists of an initial evaluation of the Portuguese AI national strategy based on the analysis of the document titled *AI PORTUGAL 2030 – PORTUGUESE NATIONAL INITIATIVE ON DIGITAL SKILLS: An Innovation and Growth Strategy to Foster Artificial Intelligence in Portugal in the European Context*.¹ Importantly, following the call from the Coordinated Plan on Artificial Intelligence, this document was released timely on June 11th 2019.² The research question is the following: how is the Portuguese AI report influenced by the EU Commission’s approach on a “trustworthy AI”? The chapter is based on documental

1 INCoDe.2030 (2019a).

2 INCoDe.2030 (2019b).

analysis³ and refers to three seminal documents on a trustworthy AI recently launched by the EU Commission: The White Paper on Artificial Intelligence;⁴ The Policy and Investment Recommendations for Trustworthy AI;⁵ and the Ethics Guidelines for Trustworthy AI.⁶ Likewise, this chapter aims to offer constructive feedback for the reviewers of the Portuguese AI national strategy.

2. The Foundations

THE PORTUGUESE AI national strategy originated from a governmental initiative established in 2016 and launched on April 3rd 2017 as the National Initiative on Digital Skills (INCoDe.2030).⁷ In the following year, on February 15th 2018, the Portuguese Council of Ministers officially approved it as a strategy for digital development aligned with the Industry 4.0 initiative, aiming to mobilise public and private resources for production of innovative knowledge and increase of economic competitiveness.⁸ According to the governmental decision, the INCoDe.2030 is structured through a coordinator, a technical secretariat, and a permanent forum for digital skills.⁹ Additionally to the general coordinator, each of the five programme axes – inclusion,¹⁰ education,¹¹ qualification,¹² specialisation,¹³ and research¹⁴ – is also led by a coordinator.¹⁵

3 An in-depth analysis would require empirical methodology – field work, large number of qualitative interviews, and selected case studies – and triangulation, which were beyond its scope. However, documental analysis provides original insights on the initial stages of these national strategies and assesses the influence of EU guidelines and recommendations.

4 European Commission (2020).

5 High Level Expert Group on Artificial Intelligence (2019b).

6 High Level Expert Group on Artificial Intelligence (2019a).

7 INCoDe.2030 (2019c).

8 Conselho de Ministros (2018).

9 Ibid.

10 Inclusion implies general access of the population to digital technologies in general.

11 The focus of education consists of the formation of youth through strengthening of digital competences in all life learning cycles.

12 Qualification means professional capacitation of the population through the necessary knowledge for integration at a labour market with growing demands for digital competences.

13 Specialisation includes the qualification of workers and creation of added value to the economy through higher offer of technical courses, undergraduate and graduate education.

14 Research focuses on the conditions for production of new knowledge and active participation in international networks and programmes of research and development.

15 Conselho de Ministros (2018).

INCoDe.2030 emerged as a general strategy for development of digital skills and not as a specialised strategy for development of artificial intelligence. All of INCoDe.2030's original goals were general – access to the internet; basic digital skills; internet use in the workplace; investment in research and development; digital literacy – and not particularly related to artificial intelligence.¹⁶ Importantly, however, the original report from the foundational moment of INCoDe.2030 in 2017 emphasised the relevance of innovative data processing and intensive use of artificial intelligence and robotics as part of their framework of reference.¹⁷ Likewise, proficient knowledge of artificial intelligence was deemed relevant at both the professional and advanced level of digital skills in the original report.¹⁸

As the AI PORTUGAL 2030 report was promoted through the coordination office of the INCoDe.2030 initiative, the Portuguese AI national strategy refers to the initial preparation of INCoDe.2030 in September 2016 as the starting point of the dynamic and evolutive process that led to the strategy to foster artificial intelligence in Portugal in the European context.¹⁹ The preparation of the AI strategy started in October 2018 and was designed within the scope of INCoDe.2030.²⁰ The preamble reproduces the five axes of INCoDe.2030 and the special AI strategy should be understood as part of the general strategy for national development of digital skills.

3. The Portuguese AI National Strategy within the European context

THE PORTUGUESE AI National Strategy was developed within the European Context. After the EU Commission decided to adopt an AI Plan in April 2018, the Portuguese government took the necessary steps to elaborate its own national plan according to the EU guidelines. Evidence of European influence on the Portuguese AI national strategy comes from the title of the report, the timeline, and its content.

¹⁶ Ibid.

¹⁷ INCoDe.2030 (2017), p. 5.

¹⁸ Ibid.

¹⁹ INCoDe.2030 (2019a).

²⁰ Ibid., p. 5.

Initially, the foreword written by the Portuguese Minister of Science, Technology, and Higher Education, Manuel Heitor, clearly endorses the ethical principles of trustworthy AI, as he defends that AI should strengthen societal robustness and be made ethical-by-design.²¹ His words clearly echo the basic elements of the ethics guidelines for a trustworthy AI: It should be lawful, ethical, and robust.²² Moreover, the Portuguese Minister also announced that the national AI strategy is fully aligned with the EU Coordinated Action Plan.²³ Likewise, the coordinator of the Portuguese plan highlights the essential importance of Human-Centered European values and the guarantee of privacy protection, safety, transparency, fairness, and inclusion within the European space.²⁴ Professor Alípio Jorge also referred to the EU Coordinated Action Plan and how it shapes the national strategies of each country by “promoting strong AI research and innovation and incorporating ethical principles by design”.²⁵

According to AI PORTUGAL 2030, an ethical committee for AI and Automation should define and deploy guidelines for ethical-by-design AI.²⁶ The strategy of designing ethical artificial intelligence seems inspired by the GDPR’s strategy of regulating privacy by design.²⁷ In Lawrence Lessig’s original proposition that “code is law,” he suggests that positive characteristics may become embedded in the technology itself.²⁸ Lessig’s influential ideas marked a strong departure from the libertarian ethos initially associated with a contemporary digital society – probably best captured through the “Declaration of Independence of Cyberspace” published online by John Perry Barlow in 1996.²⁹ With the exponential growth and pervasive use of digital technologies in

21 Ibid., p. 11.

22 High Level Expert Group on Artificial Intelligence (2019a), p. 2.

23 INCoDe.2030 (2019a), p. 11.

24 Ibid., p. 12.

25 Ibid.

26 Ibid., p. 28.

27 Voigt & Von dem Bussche (2017); Piras et al (2019); Rubinstein (2011); Rubinstein & Good (2013).

28 Lessig (2003); Lessig, (2006).

29 Barlow (1996).

our contemporary societies, several scholars examined the normativity embedded in the technology and advocated for intervention due to unfair outcomes caused by artificial intelligence and algorithmic decision-making processes.³⁰ Nowadays, the ethos of law and technology seems to be best captured by the search of a Magna Carta for the web – a project launched by Tim Berners Lee in recent years.³¹ In this context, the Portuguese call for AI to be ethical-by-design seems well aligned with such ethical ambitions.³²

Additionally, the cautionary message expressed in the European guidelines was incorporated also into the substantive content of the Portuguese strategy. For instance, in the foundational report of INCoDe.2030, a couple of years earlier, inclusion would result from general and equal access to digital technologies to the people for purposes of information, communication, and interaction.³³ The challenge related to power asymmetries was deemed to result from gender imbalance, but not from critical issues, risks, and threats related to information technology itself.³⁴ In contrast to that previous document, the AI PORTUGAL 2030 report warns that vulnerable individuals may be screened and targeted with false information and that inclusion also implies digital autonomy, safety, and privacy.³⁵ Likewise, efforts should be made to teach the population to understand the risks and threats involved; which includes rescuing young people from the “false feeling that since they are ‘digital natives’, so at ease with technology and devices, their – frail and superficial – expertise protects them from hazards and attacks”.³⁶

30 Ezrachi & Stucke (2016); Ezrachi and Stucke (2017).; O’Neil (2016); Eubanks (2018); Susskind (2018); Fortes (2020a); Fortes (2020b).

31 Sample (2018).

32 I would like to thank the curiosity of the editors and their demand for an additional reflection on the “ethical-by-design” approach of the Portuguese AI National Strategy, which led to the writing of this paragraph.

33 INCoDe (2017), p. 18.

34 Ibid.

35 INCoDe.2030 (2019a), p.14.

36 Ibid.

Moreover, the previous report did not refer to any ethical concerns related to the axis of research.³⁷ AI PORTUGAL 2030 refers to ethics in research as one of the most difficult contemporary challenges and explains that trust should be built “through data curation in order to avoid biases assuring transparency in the way judgements are made promoting accountability and explainability”.³⁸ Therefore, an analysis of the report shows the incorporation of the ethical principles of respect for human autonomy, prevention of harm, fairness, and explicability under the influence of the EU guidelines. The reference to the EU influence appears again in the next part of the report with a transition from the axis of INCoDe.2030 into the Portuguese AI policy: “this strategy is fully aligned with the Coordinated Action Plan of the EU and its Member States and it is included in the INCoDe.2030, the Portuguese initiative to foster digital skills”.³⁹ The AI PORTUGAL 2030 report also states clearly that their approach is human-centric and that people are the central element of any AI manifestation, so that the core of the Portuguese national strategy is the empowerment of human beings with the necessary knowledge, skills, and means to succeed in an AI-enabled world.⁴⁰

Importantly, the EU Ethics Guidelines on AI have influenced also the detailed vision of the strategy, as AI should promote a better society and strong ethical guidelines should protect fundamental rights of citizens and core moral values.⁴¹ Among the strategic objectives are commitments previously highlighted by the foreword of the Minister of Science, Technology and Higher Education, such as AI made

37 INCoDe (2017), pp. 20-21.

38 INCoDe.2030 (2019a), p. 15.

39 Ibid., p. 16.

40 Ibid., p. 17-18. The AI PORTUGAL 2030 report also contains a section titled “AI in the World” with a brief discussion of the international setting with references to the US, China, Canada, France, Germany, and Finland. The last paragraph of this analysis provides another reference to the European influence on the Portuguese AI strategy, as it refers to the EU declaration of cooperation on ‘Artificial Intelligence for Europe’, the nomination of the High-Level Expert Group, the production of the Coordinated Action Plan and the expectation that the strategies of the Member States should be defined by mid-2019. See pp. 22-23.

41 Ibid., p. 24.

ethical-by-design and strengthening societal robustness.⁴² Moreover, protection of core values and understanding of AI impacts depend on: definition of regulatory frameworks, guidelines of ethical-by-design AI through an ethical committee, social inclusion through technological awareness, and assessment of AI societal impacts on employment, democracy, and fairness.⁴³ The description of the planned actions also include the specific objective of guaranteeing that AI is ethically and safely applied to its various domains.⁴⁴

Current debates on the future of AI are producing concerns about, for example, transparency (promoting fairness, accountability, and the ability to explain AI systems' decisions) and computational ethics (machines that can acquire, learn, discuss, and adapt moral principles through algorithmic procedures).⁴⁵ This reference to transparency as part of the fundamental research for the future of AI evokes an ethical concern of the EU Ethics Guidelines for Trustworthy AI, but as part of a research agenda for the future instead of a present regulatory commitment. In this sense, the AI PORTUGAL 2030 report does not reproduce the same requirements recommended by the AI HLEG on the transparency of the data, the system and the business model.⁴⁶ The EU Ethics Guidelines for Trustworthy AI emphasise the importance of traceability of the decisions made by AI systems (including processes of data gathering, data labelling, and the algorithms used), which should be documented to enable identifications of the reasons for AI-decisions, to facilitate auditability, and to prevent future mistakes.⁴⁷ Additionally, explainability requires the ability to explain technical processes and related human decisions in due time and in a way that is adapted to the concerned stakeholder (e.g. layperson, regulator, researcher), including explanations on the organisational decision-making process, design

42 Ibid., p. 25.

43 Ibid., p. 28.

44 Ibid., p. 31.

45 Ibid., p. 33.

46 High Level Expert Group on Artificial Intelligence (2019a), p. 18.

47 Ibid.

choices of the system, and its rationale.⁴⁸ Furthermore, the AI HLEG considers that humans have the right to be informed that they are interacting with an AI system and the option to decide for human interaction.⁴⁹ Communication also includes the duty to provide information to practitioners and end-users on the capabilities and limitations of AI systems.⁵⁰ In comparison to this precise set of requirements for a transparent AI in the EU Ethics Guidelines for Trustworthy AI, the AI PORTUGAL 2030 report provides more general references to governmental and business transparency without the same level of detail on the requirements and the strategy for guaranteeing it.⁵¹

The national AI strategy also focuses on modernisation of public administration.⁵² Likewise, planned actions include digital inclusion and education, which are features of both the general strategy for digital skills and the specialised strategy for artificial intelligence.⁵³ Finally, planned actions also include societal challenges brought by AI to ethics and safety with responses to demands for transparency, auditability, privacy protection, and fairness.⁵⁴ Best practices and mechanisms against AI misuse are needed and legal reform will define liability related to AI decision making.⁵⁵

Overall, the Portuguese AI national strategy was clearly drafted under direct influence of the European guidelines. However, references to the three basic components – being lawful, ethical, and robust – and to the four ethical principles – respect for human autonomy, prevention of harm, fairness, and explicability – are stronger than references to the seven key requirements for the realisation of trustworthy AI. For instance, there is a clear announcement that AI should be ethical and follow the ethical principle of respecting human autonomy, but the

48 Ibid.

49 Ibid.

50 Ibid.

51 InCoDe.2030. (2019a), p. 26.

52 Ibid., p. 34.

53 Ibid., p. 36.

54 Ibid., p. 37.

55 Ibid.

Portuguese AI national strategy seems weaker in its commitment to the realisation of the requirement of human agency and oversight.

According to the EU Ethics Guidelines for Trustworthy AI, respect for human autonomy requires protection of fundamental rights, the centrality of user's autonomy, and oversight through governance mechanisms like the human-in-the-loop approach.⁵⁶ Even if there are general references to the human-centered approach in the AI PORTUGAL 2030 report, some parts of the document refer to autonomous algorithmic decision-making without a clear message around the centrality of user's autonomy, protection of fundamental rights, and human oversight.⁵⁷ AI PORTUGAL 2030 also does not condemn autonomous weapon systems – a target of both the EU Coordinated Plan on Artificial Intelligence⁵⁸ and the Ethics Guidelines for Trustworthy AI.⁵⁹ The Portuguese silence on Lethal Autonomous Weapons Systems (LAWS) obviously does not imply a misalignment with the European guidelines, but it seems a lost opportunity to declare support for the human-in-the-loop approach according to an emerging consensus among AI architects.⁶⁰

Importantly, the closing remarks in AI PORTUGAL 2030 bring an explanation on the document's life cycle and clarify that the strategy will be monitored by a Committee coordinated by the Foundation for Technology and Science and will be subject to annual reviews.⁶¹ As this report was originally released on June 11th 2019, the European

56 High Level Expert Group on Artificial Intelligence (2019a), pp. 15-16.

57 For instance, there is a paragraph on real time decision making with AI that describes how AI algorithms make autonomous decisions in many applications and must decide promptly and accurately. Autonomous algorithmic decision-making seems a major source of concern for the EU guidelines for Trustworthy AI, but the AI PORTUGAL 2030 report does not indicate a commitment to the requirement of human agency and oversight in this passage of the document (see INCoDe.2030 (2019), p. 30). There are other references to autonomous algorithmic decision-making without human oversight in the report too (see pp. 31-32). Moreover, the discussion on the future of AI research includes also autonomous AI (for driving, information systems, cybersecurity, smart cities, industry, etc) and autonomous machine learning (intelligent systems that can use machine learning autonomously) (see p. 33).

58 European Commission (2018), p. 8.

59 High Level Expert Group on Artificial Intelligence (2019a), p. 34.

60 Ford (2018), pp. 31-32, 58-61, 106-108, 138, 179, 440 and 507-508.

61 INCoDe.2030 (2019a), p. 37.

influence was restricted to the EU guidelines and did not include the Policy and Investment Recommendations for Trustworthy AI and the White Paper on Artificial Intelligence subsequently released. The annual review should update the Portuguese AI national strategy and internalise insights from these new EU documents. Likewise, the annual review could also incorporate more strongly the seven key requirements for the realisation of trustworthy AI. A few sections could also be polished and improved in the text.⁶² Importantly, however, by the time of writing of this report, there was no news about the first annual review of AI PORTUGAL 2030 on the INCoDe website.⁶³

4. Particular points of the Portuguese AI National Strategy

REGARDING POINTS OF this national strategy, this report acknowledged the large number of definitions and chose to refer to AI as “the scientific area and the suite of technologies that use programmes and physical devices to mimic advanced facets of human intelligence”.⁶⁴ The report did not explain the reasons for this broad definition, but the inspiration possibly came from Alan Turing’s imitation game mentioned in the foreword by the coordinator of the strategy.⁶⁵ In any event, this thin definition of AI provided an umbrella concept for big data processing apparently intelligent without actually exhibiting any of the properties of machine learning or pretending to achieve artificial general intelligence. It also provided examples of potential applications and contributions of AI: “autonomy, problem solving, complex planning, negotiation, reasoning, inference, decision making, diagnosis, prediction, adaptation to new situations, language understanding

62 For example, there is a reference to the creation of a technological bridge with the so-called ‘death valley’ at page 15 of the report, but the authors possibly meant the ‘silicon valley’.

63 INCoDe.2030 (2019d). Even if the document was released earlier, the website also refers to the date of November 6th 2019 as the moment of the validation and official announcement of the AI PORTUGAL 2030. Therefore, by the time of the conclusion of our report in October 2020, the first anniversary of this validation of the Portuguese AI National Strategy had not yet been reached.

64 INCoDe.2030 (2019d), p. 18.

65 Ibid., p. 12.

and generation, explanation, argumentation, visual/audio recognition, object recognition and the generation of complex artefacts”.⁶⁶ It works due to the current state-of-the-art industry and the large number of relevant big data applications that do not fit within a “thick” definition.⁶⁷

The relevance of abundant data is also mentioned at the foreword as a decisive factor for the rise of machine learning and the new golden era of AI,⁶⁸ but the strategic role of big data for the Portuguese AI national strategy is not fully developed in the report. There are generic references to a flourishing data market⁶⁹ and to the need to invest in storage, availability, and distribution of data,⁷⁰ but no discussion of the difficulties of scale faced by Portugal as a small country in comparison to China and the US. Importantly, Portugal managed this difficulty in the development of natural language processing by expanding its data and benefitting from additional sources from outside the country. However, AI PORTUGAL 2030 does not refer to this challenge of small scale, nor to the potential strategic solutions. Likewise, there are no details on the amount of funding and investments. Apparently, this is a limitation not only of this report, but also of the national strategy for digital skills in

66 Ibid., p. 18.

67 I would like to thank the reviewers for asking me an additional explanation of what I mean by a thick definition of artificial intelligence. In social sciences, the anthropologist Clifford Geertz developed the notion that concepts may be developed through a “thick description” that includes cultural context and internal meanings (see Geertz (1973); Geertz (2003); Geertz (2008); Shankman et al. (1984)). Therefore, thick definitions of Artificial Intelligence would be coined according to the particular cultural context of the field now and the internal meanings of artificial intelligence for the current architects of artificial intelligence. Such a thick definition could be based on particular techniques like Artificial Neural Networks (ANN), machine learning and the quest for general artificial intelligence that would probably exclude several applications of digital technology and processing of big data which seem intelligent, according to the thin definition used by the Portuguese AI National Strategy. Importantly, the High-Level Expert Group produced a more thick definition of Artificial Intelligence in their document A Definition of AI: Main Capabilities and Disciplines (2019c).

68 INCoDe.2030 (2019d), p. 12.

69 Ibid., pp. 24-25.

70 Ibid., p. 17.

general.⁷¹ Perhaps the Portuguese government could consider structuring and financing the AI national strategy.

On the other hand, the AI PORTUGAL 2030 report is very clear about the importance of innovation and how it should be linked to the use and development of AI. The report refers to a highly innovative Portuguese AI ecosystem and to the role of three Digital Innovation Hubs (DIHs) already established to support companies to become more productive, including SMEs.⁷² The detailed vision reveals the ambition to position Portugal as a living laboratory for experimentation of new developments and to contribute to new knowledge and development through AI research and innovation.⁷³ The report shows the expansion of Research and Development in the business sector, especially SMEs since 2015.⁷⁴ It also covers the educational system, focusing on challenges related to the performance of students⁷⁵ and linking education with digital inclusion, but with emphasis on STEM knowledge (Science, Technology, Engineering, and Mathematics), but no reference to ethics and data protection as part of the specific objectives and specific actions related to education.⁷⁶ As explained before, the AI PORTUGAL 2030 report was launched before the Policy and Investment Recommendations for a Trustworthy AI and the annual review should incorporate the important recommendations related to generating appropriate skills and education for AI.⁷⁷ This challenge is not limited to redesigning the educational systems from pre-school to higher education,⁷⁸ but especially to the development and implementation of an European Curriculum in AI which would incorporate applied ethics

71 In 2019, a High-Level Review Committee for INCoDe.2030 formed by independent reviewers Laurent Crouzet, Donatella Castelli, and Nora McGregor praised the Portuguese initiative, but made 20 recommendations to the government, including the need for more structured organization – a project manager and staff – and resources – a dedicated budget (2019), p. 10.

72 INCoDe.2030, p. 20.

73 Ibid., p. 25.

74 Ibid., 19.

75 Ibid., p. 21.

76 Ibid., p. 36.

77 High Level Expert Group on Artificial Intelligence (2019b).

78 Ibid., pp. 32-33.

to AI and ensure an interdisciplinary and multidisciplinary perspective through cooperation across technical, humanist, social sciences, legal and philosophical approaches to AI.⁷⁹ The annual review of AI PORTUGAL 2030 should consider the inclusion of a national plan for developing applied ethics to AI and to ensure such an interdisciplinary and multidisciplinary perspective as recommended by the AI HLEG.

The report is relatively laconic about the potential role of law and regulatory settings to foster trustworthy AI and contribute to technological development, but mentions regulatory frameworks as part of the strategy for developing the AI ecosystem.⁸⁰ However, there are no detailed references to concrete measures for raising consumer awareness and guaranteeing consumer redress, even if Portugal has a strong tradition in consumer law and collective redress.⁸¹ Regarding environmental protection, the report advocates the application of AI for sustainable urban transformation, sustainable energy systems, and protection of biodiversity.⁸²

Finally, AI PORTUGAL 2030 presents specific national strengths related to the attractiveness of Portugal for knowledge intensive companies and production units with potential for development of specialised AI software and high-tech devices for export.⁸³ This particular point is really important, especially in connection with the White Paper's reference to a "lighthouse centre of research and innovation for AI in Europe" that "would attract talent from all over the world due to the possibilities that it could offer".⁸⁴ The EU should seriously consider Portugal as a potential host country for such a lighthouse, not only for the concrete efforts already made to develop its digital ecosystem – including also becoming the host country for the Web Summit – but also for the symbolic importance of its historical role as a land of

⁷⁹ Ibid., p. 33.

⁸⁰ INCoDe.2030 (2019a), p. 28.

⁸¹ Ferro (2015).

⁸² INCoDe.2030 (2019a), pp. 31-32.

⁸³ Ibid., p. 27.

⁸⁴ European Commission (2020), p. 6.

navigators and discoverers.⁸⁵ European authorities should not forget that Portuguese lighthouses are currently found in China, India, and African countries as are part of the cultural heritage.⁸⁶ Likewise, as the internet remains the most visible and lucrative application of digital technology, the analogy with navigation provides a strong symbolic reference to justify the establishment of a digital lighthouse for the inter-nauts of today in the land of the navigators of yesterday.⁸⁷

5. Concluding remarks

PORTUGAL BUILT ITS national strategy for artificial intelligence on the previous governmental initiative for development of digital skills INCoDe.2030. EU guidelines influenced the AI PORTUGAL 2030 report, but the annual review should update the Portuguese AI national strategy to include insights from the Policy and Investment Recommendations for Trustworthy AI and the White Paper on Artificial Intelligence. Likewise, reviewers could incorporate more strongly the seven key requirements for the realisation of trustworthy AI. Additionally, the report could further develop the strategic role of big data, description of the amount of funding and investments, definition of specific objectives and actions related to education of ethics and data protection, and explanation of the role of law, regulation, and redress to development of a trustworthy AI. In any event, Portugal looks attractive for companies and could eventually become the host country for a European digital lighthouse.

85 Likewise, the history of modern navigation and discoveries was also a period of colonial expansion, political imperialism, ethnic and social domination, transatlantic slavery and servitude. This critical reflection on the Portuguese historical legacy and the role of European settlers and seafarers around the globe actually also provides a strong warning about the need for ethical digital navigation, exploration, and commerce, so that our contemporary digital societies do not reproduce the unfair practices of the previous ones.

86 Prodigious examples are the Guia Lighthouse in Macao, the Aguada Lighthouse in Goa, and the Dona Maria Pia Lighthouse in Cape Verde.

87 Falcão (2017).

References

- Barlow, J. P. (1996). A Declaration of the Independence of Cyberspace. Electronic Frontier Foundation. <https://www.eff.org/pt-br/cyberspace-independence>
- Cathy, O. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
- Conselho de Ministros. (2018). Resolução do Conselho de Ministros n.º 26/2018. Diário Da República, 1(48–8), 1207–1209.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.
- European Commission. (2018). Communication Coordinated Plan on Artificial Intelligence (COM(2018) 795 final). European Commission. https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en
- European Commission. (2020). White Paper On Artificial Intelligence—A European approach to excellence and trust (COM(2020) 65 final; p. 27). https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
- Ezrachi, A., & Stucke, M. E. (2016). *Virtual Competition: The Promise and Perils of the Algorithm-driven Economy*. Harvard University Press.
- Ezrachi, A., & Stucke, M. E. (2017). *Artificial Intelligence & Collusion: When Computers Inhibit Competition*. University of Illinois Law Review, 2017. <https://papers.ssrn.com/abstract=2591874>
- Falcão, J. (2017). *Mundus Novus: The Construction of a Free Flow of Information from the Navigators of Yesterday to the Internauts of Today*. In P. R. B. Fortes, L. Boratti, A. Palacios Lleras, & T. Gerald Daly (Eds.), *Law and Policy in Latin America: Transforming Courts, Institutions, and Rights*. Palgrave Macmillan UK.
- Ford, M. (2018). *Architects of Intelligence: The truth about AI from the people building it* (Packt Publishing).
- Fortes, P. R. B. (2020a). O consumidor contemporâneo no Show de Truman: A geodiscriminação digital como prática ilícita no direito brasileiro. *Revista de Direito do Consumidor*. <https://revistadedireitodoconsumidor.emnuvens.com.br/rdc/article/view/1176>
- Fortes, P. R. B. (2020b). Paths to Digital Justice: Judicial Robots, Algorithmic Decision-Making, and Due Process. *Asian Journal of Law and Society*, 1, 1–17. <https://doi.org/10.1017/als.2020.12>
- Geertz, C. (1973). *The Interpretation Of Cultures*. Basic Books.
- Geertz, C. (2003). Thick Description: Toward an Interpretive Theory of Culture. In Y. S. Lincoln & N. K. Denzin (Eds.), *Turning Points in Qualitative Research: Tying Knots in a Handkerchief*. AltaMira Press, U.S.
- Geertz, C. (2008). *Local Knowledge: Further Essays in Interpretive Anthropology*. Basic Books.
- High-Level Expert Group on AI. (2019a). Ethics guidelines for trustworthy AI (p. 41). Set up by the European Commission. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- High-Level Expert Group on AI. (2019b). Policy and investment recommendations for trustworthy Artificial Intelligence (p. 52). Set up by the European Commission. <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>
- High-Level Expert Group on AI. (2019c). A Definition of AI: Main Capabilities and Disciplines (p. 9). European Commission. <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

High-Level Review Committee. (2019). Review of the National Digital Competences Initiative, Portugal INCoDe.2030 (p. 15). https://www.incode2030.gov.pt/sites/default/files/2019-incode2030-hlrc-finalreport_6.pdf

INCoDe.2030. (2017). Iniciativa Nacional de Competências Digitais e.2030: Um Programa Integrado de Competências Digitais para Portugal. https://www.incode2030.gov.pt/sites/default/files/uploads/attachments/incode2030_final_28mar17.pdf

INCoDe.2030. (2019a). AI PORTUGAL 2030. Portugal INCoDe.2030. <https://www.incode2030.gov.pt/en/ai-portugal-2030>

INCoDe.2030. (2019b). Portugal lança a Estratégia Nacional para a Inteligência Artificial. Portugal INCoDe.2030. <https://www.incode2030.gov.pt/destaque/portugal-lanca-estrategia-nacional-para-inteligencia-artificial>

INCoDe.2030. (2019c). Equipa de Recursos e Tecnologias Educativas. <https://erte.dge.mec.pt/incode2030>

INCoDe.2030. (2019d). Portugal launches the National Strategy for Artificial Intelligence. Portugal INCoDe.2030. <https://www.incode2030.gov.pt/en/featured/portugal-launches-national-strategy-artificial-intelligence>

Lessig, L. (2003). Law Regulating Code Regulating Law. *Loy. U. Chi. L. J.*, 1. <https://dash.harvard.edu/bitstream/handle/1/12912675/Law%20Regulating%20Code%20Regulating%20Law.pdf>

Lessig, L. (2006). *Code: And Other Laws of Cyberspace*, Version 2.0 (2nd ed.). Basic Books.

Piras, L., Al-Obeidallah, M. G., Praitano, A., Tsohou, A., Mouratidis, H., Crespo, B. G.-N., Bernard, J. B., Fiorani, M., Magkos, E., Sanz, A. C., Pavlidis, M., D'Addario, R., & Zorzino, G. G. (2019). DEFEND Architecture: A Privacy by Design Platform for GDPR Compliance. Trust, Privacy and Security in Digital Business - 16th International Conference, TrustBus 2019, Proceedings, 78–93. https://doi.org/10.1007/978-3-030-27813-7_6

Rubinstein, I. (2011). Regulating Privacy by Design. *Berkeley Technology Law Journal*, 26, 1409.

Rubinstein, I., & Good, N. (2013). Privacy by Design: A Counterfactual Analysis of Google and Facebook Privacy Incidents. *Berkeley Technology Law Journal*, 28, 1333. <https://doi.org/10.2139/ssrn.2128146>

Sample, I. (2018, November 5). Tim Berners-Lee launches campaign to save the web from abuse. *The Guardian*. <https://www.theguardian.com/technology/2018/nov/05/tim-berners-lee-launches-campaign-to-save-the-web-from-abuse>

Shankman, P., Ágh, A., Bourguignon, E., Brintnall, D. E., Cole, J. R., Connor, L., Darnell, R., De Ruijter, A., Dutton, D., Fabian, J., Farrer, C. R., Fisher, A. D., Howe, L. E. A., Richardson, M., Ridington, R., & Wilk, S. (1984). The Thick and the Thin: On the Interpretive Theoretical Program of Clifford Geertz [and Comments and Reply]. *Current Anthropology*, 25(3), 261–280.

Sousa Ferro, M. (2015). Collective Redress: Will Portugal Show the Way? *Journal of European Competition Law & Practice*, 6, 299–300.

Susskind, J. (2018). *Future Politics: Living Together in a World Transformed by Tech*. Oxford University Press.

Voigt, P., & Bussche, A. von dem. (2017). *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-57959-7>

AI policy in Poland

Ethical considerations already at the core

Kasia Söderlund

LLM, PhD student in Technology and Society, LTH, Lund University, Sweden, and a member of the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS).

Summary

THE CHAPTER OUTLINES the main stages of development of AI strategy in Poland, focusing on three significant documents: *Considerations for AI Strategy in Poland* (November 2018), *Policy for Development of AI in Poland for the years 2019-2027* (August 2019) and the draft *Policy for Development of AI in Poland from year 2020* (September 2020). At the moment of writing, the latter document has been adopted by the Polish Committee of the Council of Ministers for Digitisation and is still subject to the ongoing legislative process. It should therefore be noted that the final shape of the Polish AI strategy may differ to some extent from the one analysed in this chapter. The report was prepared through analysis and comparison of the above documents, with a few references to preparatory documents and official governmental websites.

What follows is a deeper analysis of the second draft of the AI national strategy, *Policy for Development of AI in Poland from year 2020*, with regard to how the following points of interests were interpreted and incorporated: definition, funding and investments, applications, data, the country's identified characteristics and strengths, innovation, law, education, consumers, SMEs, environment and inequality. The strategy sets out objectives which Poland is planning to achieve in the short-term (until 2023), medium-term (until 2027), and long-term (beyond 2027). It is divided into five subject areas: society, innovative companies, science, education, international cooperation and public sector, and consists mainly of action points on different levels of detail.

The analysis in this chapter of the drafted AI strategy shows that the ethical component – especially the EU's and OECD's concepts of trustworthy AI – played an important role in shaping the strategy. It is also noted that ethical principles were already reflected to a high degree in the first document, *Considerations for AI Strategy in Poland*, which suggests that the ethical dimension of AI has been taken seriously by Polish policymakers throughout the legislative process.

1. Introduction

IN DECEMBER 2018, at the time when the European Commission (hereafter just Commission) called for the EU member states to establish their national AI strategies by mid-2019, Poland already had in place a solid base for development of its own strategy. A comprehensive document, *Considerations for AI Strategy in Poland*, adopted in November 2018,¹ was the first step in the process of development of the AI policy. The first draft of the strategy was the document that followed, *Policy for Development of AI in Poland for the years 2019–2027*, released for public consultations in August 2019.² After many months of inter-ministerial, expert and further social consultations, the Committee of the Council

1 Ministerstwo Cyfryzacji (2018).

2 Ministerstwo Cyfryzacji (2019).

of Ministers for Digitization³ published in September 2020 the second draft of the AI strategy, *Policy for development of AI in Poland* (the ‘Draft’ or the ‘Policy draft’).⁴ At the moment of writing of this report, however, the document is still subject to an ongoing legislative process and the exact date of approval of the final AI Policy is unknown. However, the urgency for adopting the AI Policy is recognised by the government, especially because of the special procedure in which the Policy draft is being considered.⁵

What follows is a brief description of the two documents preceding the Draft, with a particular focus on the issue of the ethical considerations and the concept of Trustworthy AI. Thereafter, it will be examined how the following points of interests were interpreted and incorporated in the Policy draft: the definition of AI, funding and investments, applications, data, the country’s identified characteristics and strengths, innovation, law, education, consumers, SMEs, environment and inequality.

2. Considerations for AI Strategy in Poland, November 2018

POLAND’S PREPARATORY ACTIVITIES for establishing its AI strategy were initiated in parallel with the works on the European AI strategy at the Commission level in 2018. The Ministry of Digitisation invited a group of more than 180 experts from the industry, public sector, academia and civil society, to examine the areas of data-driven economy, financing, education, law and ethics, in order to pave the way for development of AI strategy in Poland. The result of the consultations was a 250-page-long document, published in November 2018, which would later serve the Polish government as a foundation for the first draft of the AI policy.

3 Committee of the Council of Ministers for Digitization (pol. Komitet Rady Ministrów ds. Cyfryzacji), an auxiliary body of the Council of Ministers and the Prime Minister established to ensure the coordination of the implementation of IT projects of the government administration and the preparation of government documents related to digitisation.

4 Ministerstwo Cyfryzacji (2020a).

5 Ministerstwo Cyfryzacji (2020b), p. 13.

It is noteworthy that the experts dedicated a significant part of the *Considerations* to deal with the ethical component of AI, and recommended a list of ethical principles to be included in the strategy. Since the *Considerations* were published before the AI HLEG's Ethics Guidelines,⁶ the suggested ethical principles in the *Considerations* were slightly different from the ones listed in the Ethics Guidelines. Similarities included the general approach to the ethical principles as stemming from the fundamental and democratic rights, with the right to human dignity at the core. Apart from the principles of fairness, explicability, transparency, digital inclusion, etc., the list of ethical principles to be protected also included the special protection of vulnerable persons, the issue of AI's influence on the labour market, and a proposal to limit AI's application scope for some cases of automated decision-making in the public sector.

In general, the document presents a very useful, in-depth analysis of the AI infrastructure with many recommendations coming from the experts, and it therefore served as a well-grounded basis for drafting the first version of the AI Policy.

3. Policy for Development of AI in Poland for the years 2019–2027, August 2019

THE FIRST DRAFT of the AI strategy, *Policy for Development of AI in Poland for the years 2019–2027*,⁷ was published in August 2019 and was open for public consultations until September 2019. This time, since the AI HLEG's Ethics Guidelines had already been released, the ethical principles and requirements of the Guidelines were clearly reflected in the Policy draft.

Similarly as in the *Considerations*, the starting point for the reasoning regarding the ethical element in this document was the absolute right to human dignity, recognised as the foundation of the axiological order

⁶ European Commission (2019a).

⁷ Ministerstwo Cyfryzacji (2019).

of the state.⁸ We read further that since the ethical problems related to application of AI systems are expected to only increase in the future, it is of the highest importance to maintain, or enhance, trust, and to set up an appropriate environment for design, development and use of AI technologies, which would be worthy of human trust.⁹ It is noted that the issue of trust in AI should be understood broadly, and encompass not only algorithms and other AI technologies, but also take into account the wider, consequential context of their application within societies.

Subsequently, the notion of the Human-Centric Approach on AI is invoked, along with UNESCO's framework of a *Global Ethics Code* for AI. However, also concepts like the EU's *Trustworthy AI* and the OECD's *Stewardship of Trustworthy AI* seem to have been given special, nearly equally important significance, as they are considered to be the initiatives that Poland "supports, and their recommendations adopts as its own for the purpose of this document".¹⁰ It is emphasised that the ethical policy of AI in Poland should be developed in line with the EU's concept of *Trustworthy AI*.¹¹ Both the Ethics Guidelines and *Trustworthy AI* are given more explanation in Annex 2, labelled "Ethical dimension," where large parts of these are incorporated. For example, the tables concerning the framework for *Trustworthy AI*, interrelationships between the seven requirements and the system's life cycle, are translated and reproduced, and the components, ethical principles, and requirements for *Trustworthy AI* are explained. Reference is also made to the *Trustworthy AI Assessment List*, as a checklist that would help verify the application of each of the key ethical requirements set out in the Guidelines.

8 Ibid., p. 38.

9 Ibid., p. 40.

10 Ibid., p. 40.

11 Ibid., p. 41.

4. The draft of the Policy for Development of AI in Poland from year 2020, September 2020

SO FAR, THE *Policy for Development of AI in Poland from year 2020* has not yet been adopted as an official Polish AI strategy, although the final version of the Policy was initially planned to be presented in the first quarter of 2020. It can be supposed, however, that the role of the coronavirus outbreak in 2020 cannot be disregarded in shifting of the governmental priorities and as a result, contributing to the delay.

Once the final Policy is adopted, it will be Poland's response to the Commission's *Coordinated Plan on Artificial Intelligence*,¹² and it will also implement a few important national programmes and international strategies concerning AI, innovation and digitisation strategies, such as *Ethics Guidelines for Trustworthy AI*,³³ proposed Policy and Investment Recommendations, and OECD recommendations. The Policy draft is a result of concerted efforts of governmental institutions, public sector, industry, civil society and academia, in order to build a comprehensive and dynamic national AI strategy, and to spur the development of AI technologies in Poland.

The overarching purpose of the Policy will be to support society, companies, scientists and the public sector in taking advantage of the opportunities created by AI technologies and, at the same time, to ensure protection of human dignity and fair competition in the global context. The resulting document is over 60 pages long, divided into six sections:

- ▶ AI and society – setting out the list of activities and objectives to develop Poland's data-based economy, and to make the Polish society aware of the need to constantly improve its digital competences;
- ▶ AI and innovative companies – aiming at supporting Polish AI enterprises, including financing mechanisms for their development,

12 European Commission (2018), see also chapter 1 in this volume.

cooperation between start-ups and the government, and creation of regulatory “sandboxes”;

- ▶ AI and science – support for the Polish scientific and research community, foundation of AI doctoral positions, grants for researchers and other activities aimed at preparing an expert AI workforce, including a framework for the ethical and safe use of this technology;
- ▶ AI and education – actions on the first-, secondary-, and third-cycle educational level, programs for people at risk of losing their jobs as a result of the digital transformation, educational grants;
- ▶ AI and international cooperation – activities on the international arena with the objective of supporting the promotion of Polish businesses in the field of AI and the development of AI technologies with respect for human dignity and human rights, in line with EU and OECD standards;
- ▶ AI and the public sector – aiming at supporting the public sector in, among other objectives, facilitating governmental public procurement for AI solutions, improved coordination of AI development, opening and sharing as much public data as possible with citizens and companies.

Each of the above sections are further divided into subsections listing specific actions and objectives to be reached in the respective timeframes: short-term perspectives (until 2023), middle-term perspectives (until 2027) and long-term perspectives (after 2027).

In general, the Draft is a concise document, describing the digital revolution as a window of opportunity for Poland’s development leap. It contains very specific sets of actions and objectives in high level of detail, while at the same time remaining adaptable and open for the possible adjustments.

4.1. “Trustworthy AI”

The concept of Trustworthy AI has been incorporated into the Policy draft as one of the main pillars of the AI ecosystem.¹³ As a succinct

13 Ministerstwo Cyfryzacji (2020a), e.g., p. 59, “[t]he Polish ecosystem operates in relation to the international and legal dimension, technical and organizational standards, and above all, a human functioning in society and the environment on the basis of ethical principles for Trustworthy Artificial Intelligence.”

document, it does not explain in detail the concept of Trustworthy AI. Instead, it refers the reader to the Ethics Guidelines, and emphasises that development of AI should be underpinned by the European approach to the concept.¹⁴ Only the seven requirements for realisation of Trustworthy AI are explicitly listed in Annex 1, “Definition of AI.”¹⁵

However, apart from the AI HLEG’s Ethics Guidelines, other initiatives aimed at the development of ethical AI on other fora are mentioned,¹⁶ such as the UNESCO’s *Global Ethics Code for AI*, OECD’s *Stewardship of Trustworthy AI* and the Council of Europe’s works towards recommendations for AI respecting human rights, the rule of law and democracy. Poland declares its willingness to take part in these works, and it is even claimed that Poland is one of the most active supporters of ethical use of data, in accordance with the idea of Trustworthy AI.¹⁷

Additionally, a Polish *Ethics AI Code* is planned to be developed, which would be based on Article 30 of the Polish Constitution (the right to human dignity) and the Charter of Fundamental Right of the EU.

4.2. AI definition

The problem of defining AI for the purposes of the strategy is laid out separately in Annex 1. As is usually the case with such definitions, the starting point is a reflection on the challenges in defining AI, and that no agreement as for the legal definition of AI has been reached thus far. Nonetheless, the policymakers approached the task by describing the phenomenon of AI, pointing to the main features of AI systems and explaining their essential components.

The section is introduced by an explanation of AI in a “collective” way, by the Polish think-tank Sobieski Institute,¹⁸ as an area including e.g. robotics, machine learning, creation of intelligent behaviour models

¹⁴ Ministerstwo Cyfryzacji (2020a), p. 21.

¹⁵ *Ibid.*, p. 58.

¹⁶ *Ibid.*, pp.53–54.

¹⁷ *Ibid.*, p. 51.

¹⁸ Michałowski (2018).

and computer programs simulating these behaviours.¹⁹ Later, the narrative zooms out onto the international level. It is observed that a certain consensus as for the understanding of AI has been resolved, although it captures the definition of AI from its operational side, based on the technical development of the intelligent agent model. Here, the Draft refers to the OECD's definition of AI systems in *Recommendation of the Council on Artificial Intelligence*,²⁰ as well as to the AI HLEG's explanation of AI in "A definition of AI: Main capabilities and scientific disciplines of AI".²¹ Both definitions are partially reproduced in the text of the Annex, and as mentioned above, reference is made to the seven key requirements of Trustworthy AI.

4.3. Funding and investments

Since the number of large private companies in Poland is relatively small, the responsibility of financing AI projects is considered to rest mainly on the public sector and state-owned companies. A list of the 20 main programmes financing implementation and development of innovative projects was incorporated into the text of the document. These include programmes such as support of VC funds investing in innovative technologies, green technology initiatives (lowering carbon emissions and reducing the degrading effect on the environment), investment in start-ups (including some of the high-risk profile), scientific research in AI and machine learning, investments in Polish as well as European innovation hubs, and encouraging medium- and large companies to support tech accelerators.

Some of the interesting examples of tools intended to increase the level of innovation, including AI solutions, are the following:

¹⁹ Ibid., p. 12.

²⁰ "A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments" (OECD, 2020: 7).

²¹ European Commission (2019b).

- ▶ including the requirement of AI solutions in specifications of strategic investments financed from public funds,
- ▶ adaptation of currently used tax mechanisms for enterprises investing in innovative solutions,
- ▶ creating the possibility of testing ground, water and air vehicles and autonomous ships in designated zones (road sections, public waters, smaller towns, districts of larger cities) and under certain conditions.

The programmes listed in the Draft amount to around 2 billion EUR (9180 million PLN). It is important to note, however, that the government's investments in AI projects are included into the broader financing category of innovation. The programmes are meant to be periodically evaluated as for their suitability, and special consideration will be taken as to whether sufficient funds are allocated for investments in AI.

It has been pointed out by one of the Ministries, however, that the financing framework should be more detailed, and contain a specified schedule of activities with assigned entities responsible for their implementation.²² The government's response to this concern was that a detailed action plan will be developed after the adoption of the main text of the *AI Policy*, in order to uphold its flexible and dynamic character.²³

4.4. Applications

Some concrete examples of AI applications mentioned in the analysed document include: energy, environmental protection, water saving, health and senior care, tools predicting the development of the epidemiological situation, communication and housing infrastructure, protection of the external border of the EU, autonomous vehicles, cinematography (digital reconstruction, automatic film translation, dubbing preparation).

Moreover, certain sectors of the economy are prioritised and expected to gain the most from the development of AI technologies. These

²² Ministerstwo Cyfryzacji (2020c), com. 49.

²³ Ibid

sectors include state administration, construction (in particular smart building), cybersecurity, energy, trade and marketing, medicine, industry, agriculture, transport and logistics. The Policy draft estimates that benefits of implementing AI in the prioritised sectors are equivalent to around 2.65 percent of the Polish GDP.

4.5. Data

The significance of data, and its fundamental role in the digital transformation of societies, is raised at the very beginning of the document.²⁴ It is acknowledged that acquiring, collecting, analysing, processing, using data and the constant development of AI algorithms are crucial for data-driven economies, and in order to enter the era of AI, public and commercial services must be deeply saturated with data.

4.6. Country's strengths

Despite Poland not ranking at the top of ICT ratings,²⁵ the Policy draft highlights a few areas that Poland might focus on in order to boost its potential. Some factors are specifically pointed out as likely to contribute to the Polish development in AI, for example:

- ▶ Students' high results in mathematics and natural sciences in OECD's PISA assessments, in which Polish 15-year-old students were ranked 10th in 2018.²⁶
- ▶ According to a Global Creativity Index compiled by management studies specialists,²⁷ approximately 33 percent of the Polish population could be classified as belonging to the so-called creative class – comparable to The United States, which has a similar figure (33 percent) – making the country comparable also regarding niche applications of AI and attendant research.
- ▶ A thriving computer games industry, which is a domain of the tech sector that is closely related to AI.

²⁴ *Policy for Development of AI in Poland*, p.8.

²⁵ OECD Data (2020).

²⁶ OECD (2018).

²⁷ Florida et al. (2015), p. 35. For comparison, the report lists the figure for Sweden at 45 percent, the Netherlands, Canada, and the United Kingdom all at 44 percent.

Another specific Polish advantage, although not listed in the Draft, is the population's relatively high proficiency in English as a second language. In the EF English Proficiency Index,²⁸ Poland took the 9th global position, with a score comparable to the Nordics and Germany.

4.7. Innovation

The issues of innovation and AI technologies are to a great extent considered jointly. As already touched upon in the Funding and investments section, this is also reflected in the way AI projects are designated to be financed. The 20 programmes listed in the Policy draft, considered for financing AI projects, are intended to be included also in the broader context of investments allocated for innovation. However, the programmes are sufficiently adaptable and open for a wide range of projects, so effectively each of them could finance an AI project within its scope of application. For example, a programme BRIdge Alfa is designed for innovative ideas that are in the initial stage, where the risk of investment failure is the greatest, but can be verified at a relatively low cost. Other programmes target exclusively AI solutions, such as INFOSTRATEG, supporting research on selected machine learning problems that have the potential of development on a larger scale.

4.8. Law

The approach to the regulatory settings is two-fold. On the one hand, the wider context of the EU and international law is considered, especially with regards to the Ethics Guidelines. On the other hand, the document highlights the need for monitoring and evaluating the national legislation.²⁹ For example, the policymakers pointed to the problem of unequal access to AI solutions and submitted certain action points in this regard, or defined more specific tools, including enabling the exploration of the Polish language by addressing the issue of copyright protection in texts.

Further, in Annex 2 to the Policy draft, an "AI Ecosystem" is described as

²⁸ World Economic Forum (2016).

²⁹ Ministerstwo Cyfryzacji (2020a), p.21.

a horizontal environment covering four dimensions: international, ethical, legal, as well as technical and organisational, intended to initiate and support activities undertaken by AI stakeholders. In order to implement the objectives of the future policy, the policymakers appointed an AI Legislative Team, whose task will be to address legal and ethical challenges supporting the implementation of the AI policy.³⁰ Within the legal dimension, what is specifically addressed, among others, are issues of legal definition of AI, ownership of personal data and their portability, protection of business secrets, Intellectual Property, responsibility for damages caused by AI, and support for procurement specifications for AI solutions.

4.9. Education

As regards to the areas of education and science, both sections contain many ambitious action points. However, the objectives connected with ethical and sociological aspects in AI were incorporated in the section concerning AI and society. They include, among other tools, analysis of the ethical effects of AI through research in the form of scientific grants, competitions and other financial instruments, as well as initiating the path for grants in research in the field of transparency of AI algorithms applications.³¹

More generally, on top of the above actions, the significance of including the ethical dimension in AI and science was emphasised in the document in the following way:

The key to the implementation of the AI Policy is treating this discipline as multidimensional – taking into account, apart from engineering, also humanities and social sciences, which are important for determining the subsequent framework for the use of AI in social and economic life, and in particular ethical frameworks. Only in this way it will be possible to ensure that AI is used in a manner acceptable to society.³²

³⁰ Ibid., p. 21.

³¹ Ibid., p. 22.

³² Ibid., p. 34.

This indicates a sensitivity towards subject areas other than natural sciences (engineering, economics, etc.) which adds to the image of the policymakers being genuinely concerned about human values.

4.10. Consumers

Although the Policy draft does not specifically address the issue of how consumers are to be protected with regard to the rapid development of AI technologies, many action points – by the same token serving consumers – can be found in the section on AI and society. One of the long-term goals listed therein is that Poles are aware of the opportunities and threats caused by the development of modern technologies.³³ Another goal highlights the importance of building social trust and readiness to use AI solutions combined with democratisation of access to AI. This is intended to be achieved by promoting knowledge about AI and its impact on society through the media, working against disinformation and false information about the functioning of AI, as well as campaigns to raise the awareness of the public and companies on how to handle data – in particular in the context of the use of complex algorithms.³⁴

4.11. SMEs

Providing support for small and medium enterprises (SMEs) is inherent to the Policy draft through many programmes, as enumerated in the budget for innovation funding (listed in the section “Funding and investments”). The SpeedUp Energy Innovation and EEC Magenta programmes, directed at start-up technology companies in the growth and/or expansion phase, can serve as good examples. However, there are also other ways in which the support for SMEs is intended to be realised, for example by analysis and elimination of legislative barriers and administrative burdens for new enterprises using AI.³⁵

However, the document does not stipulate that SMEs and start-ups would be specifically assisted in the process of development for

³³ Ministerstwo Cyfryzacji (2020a), p. 25.

³⁴ Ibid., p. 23.

³⁵ Ibid., p. 24.

trustworthy AI. The responsibility to include ethical standards in all AI systems was emphasised in the Policy draft, regardless of the size of the companies.

4.12. Environment

The Draft stresses the impact that AI will have on energy, climate and environment. One of the action points in AI and society is to use AI solutions to monitor and improve Poland's environment. It is also noteworthy that one of the largest programmes included in the budget for innovation financing (2500 million PLN, an equivalent of ca. 547 million EUR) is the programme *New Energy*, supporting entrepreneurs and municipalities in the implementation of innovative technologies in the energy sector, including in the areas of “smart cities” and “self-sufficient energy clusters”. Elsewhere, the Policy draft indicates that AI technologies:

...will allow, among others, to integrate and stabilize the operation of distributed generation and renewable energy sources, effective management of electricity consumption, and by increasing the flexibility and controllability of the power system, [...] increase the level of reliability of supplies and the quality of electricity delivered to consumers.³⁶

4.13. Inequality

The problem concerning inequality of access to AI and data was already noted in the *Considerations for the AI Policy*, in which the authors recommended addressing the issue. One of the documents' objectives is to facilitate development of AI solutions and access to data regardless of the size of the entity.³⁷ Some of the tools include a “Digital Administration Sandbox,” an “Open Data portal,” digital repositories created in the cultural sector, as well as commercial and academic solutions based on open data.

³⁶ Ibid., p. 24.

³⁷ Ibid., p. 24.

5. Conclusions

THE ANALYSIS OF the *Considerations for AI Strategy in Poland*, *Policy for Development of AI in Poland for the years 2019–2020*, and the draft of the final AI strategy *Policy for Development of AI in Poland from year 2020*, suggests that the ethical element is one of the central components in all the documents, and that both versions of the Policy have been strongly influenced by the EU and OECD concept of trustworthy AI.

The published draft of the Polish AI strategy in September 2020 may still be subject to some changes, but this analysis of the document shows that the Polish policymakers seem to have taken the Commission's recommendations seriously, and have incorporated the ideas of trustworthy AI at the core of the Policy draft. Although the explanation concerning the very concept of Trustworthy AI has been done in a rather rudimentary way, considering the concise character of the document, the general reference to the concept seems to play a pragmatic role. It is also notable that many of the ideas concerning the ethical dimension of AI were already put forward in the *Considerations for AI Strategy in Poland*, i.e. before the Commission's invitation to prepare the strategies and before the Ethics Guidelines were published. Poland has also expressed a strong interest in taking part in the further works on Ethical AI at the European and international level.

The Policy draft includes very specific sets of actions and objectives in a high level of detail, while at the same time remaining flexible. This is reflected in the financing framework of the project, by making it possible to adapt to the changing and specific needs of the Polish AI development in the upcoming years. It has been clarified by the government in the comments to the Policy draft that an additional, more detailed action plan will be developed after the adoption of the main text of the AI Policy, due to the fact that introducing a change to the Policy should not involve repeating the legislative path.³⁸

38 Ministerstwo Cyfryzacji (2020c), com. 49.

We should hope that the Policy will be officially adopted very soon, while keeping in mind that establishing a national AI strategy is only the beginning of the journey. Even the best policy on paper can be executed poorly, and the other way around. The Policy appears to lay out an ambitious, promising strategy, and it remains to be seen to what extent it will be realised.

References

- European Commission (2018). *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Coordinated Plan on Artificial Intelligence* (COM(2018) 795 final). Brussels, December 7, available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:795:FIN>, accessed 28th September 2020
- European Commission (2019a). *Ethics guidelines for trustworthy AI*. High-Level Expert Group on AI, April 8, available at <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, accessed 16th October 2020.
- European Commission (2019b). “A definition of Artificial Intelligence: main capabilities and scientific disciplines,” available at <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>, accessed 16th October 2020.
- Florida, R., Mellander, C., & King, K. (2015). *The Global Creativity Index 2015*. July. Toronto: Martin Prosperity Institute, Rotman School of Management, University of Toronto. Available at https://ec.europa.eu/futurium/en/system/files/ged/31-most_creative_countries_-_global_creativity_index_2015_-_canadian_mgt_school.pdf, accessed 10 October 2020
- Michałowski, B. (2018). “Internet of Things (IoT) I Artificial Intelligence (AI) w Polsce.” Report, Instytut Sobieskiego, March 21, available at <https://sobieski.org.pl/iot-i-ai-w-polsce/>, accessed 20th September 2020
- Ministerstwo Cyfryzacji (2018). *Considerations for AI Strategy in Poland* (pol. Założenia do strategii AI w Polsce) of 9 November 2018, available at <https://www.gov.pl/web/cyfryzacja/ai>, accessed 8th October 2020
- Ministerstwo Cyfryzacji (2019). *Policy for Development of AI in Poland for the years 2019–2027* (pol. Polityka Rozwoju Sztucznej Inteligencji w Polsce na lata 2019–2027) of 21 August 2019, available at <https://www.gov.pl/web/cyfryzacja/ai>, accessed 8th October 2020
- Ministerstwo Cyfryzacji (2020a). *Policy for Development of AI in Poland from year 2020* (pol. Polityka rozwoju sztucznej inteligencji w Polsce od roku 2020) of 14 September 2020, available at <https://mc.bip.gov.pl/projekty-aktow-prawnych-mc/projekt-uchwaly-rady-ministrow-w-sprawie-ustanowienia-polityki-dla-rozwoju-sztucznej-inteligencji-w-polsce-od-roku-2020.html> accessed 5th October 2020
- Ministerstwo Cyfryzacji (2020b). Justification to the Resolution of the Council of Ministers on the establishment of the “Policy for the development of Artificial Intelligence in Poland from 2020” (pol. Uzasadnienie do Uchwały Rady Ministrów w sprawie ustanowienia „Polityki dla rozwoju Sztucznej Inteligencji w Polsce od roku 2020”), available at <https://mc.bip.gov.pl/projekty-aktow-prawnych-mc/projekt-uchwaly-rady-ministrow-w-sprawie-ustanowienia-polityki-dla-rozwoju-sztucznej-inteligencji-w-polsce-od-roku-2020.html>, accessed 5th October 2020
- Ministerstwo Cyfryzacji (2020c). Table of comments to the Resolution of the Council of Ministers on the establishment of “Policy for the development of Artificial Intelligence in Poland from 2020” (pol. Tabela uwag do Projektu uchwały Rady Ministrów w sprawie ustanowienia „Polityki dla rozwoju Sztucznej Inteligencji w Polsce od roku 2020”), available at <https://mc.bip.gov.pl/projekty-aktow-prawnych-mc/projekt-uchwaly-rady-ministrow-w-sprawie-ustanowienia-polityki-dla-rozwoju-sztucznej-inteligencji-w-polsce-od-roku-2020.html>, accessed 5th October 2020

OECD (2018). OECD's 2018 PISA assessments, available at https://www.oecd.org/pisa/PISA-results_ENGLISH.png, accessed 11 October 2020

OECD (2020). *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449. Available at <https://legalinstruments.oecd.org>, accessed 16th October 2020.

OECD Data (2020). OECD's ICT indicators for Poland, available at <https://data.oecd.org/poland.htm#profile-innovationandtechnology>, accessed 5 October 2020

World Economic Forum (2016). "The English Proficiency Index 2016," November 15. Available at <https://www.weforum.org/agenda/2016/11/which-countries-are-best-at-english-as-a-second-language-4d24c8c8-6cf6-4067-a753-4c82b4bc865b/>, accessed 10 October 2020

CHAPTER 4.

AI policy in Norway

Looking to the future and harmonised with the EU

Frans af Malmborg

Frans af Malmborg is a PhD student in Public Administration at the Department of Political Science and Management at the University of Agder, Norway. His research concerns Artificial Intelligence Policy within the European Union. Frans holds an MA in Public Administration from the University of Gothenburg and a one-year MA in Social Work from Malmö University.

Summary

THE AIM OF this chapter is to analyse the extent to which the Norwegian National Strategy for Artificial Intelligence (AI) has been influenced by the European Commission's approach to AI as articulated through the Commission's High Level Expert Group on Artificial Intelligence (AI HLEG) as well as the Commissions coordinated plan on AI from 2018.¹ Drawing on an in-depth systematic analysis of the strategy as well as interviews with key government officials, we conclude that the national strategy largely follows the European approach to AI. This is the case in

¹ A month after the publication of the Norwegian National Strategy in January the EU Commission published a White Paper on AI in February. Due to the White Paper being published after the Strategy, this chapter does not consider the White Paper as a potential influential force on the national strategy.

terms of general narrative; as a tech-optimist and future oriented strategy focusing on the potentials of AI technology situated in a Norwegian context. The strategy also draws largely on the Commission's approach concerning the definition and conceptualisation of AI technologies and adopts the policy-construct of "Trustworthy AI", dedicating an entire section on how the principles promoted by the AI HLEG is going to be implemented in Norway. Word queries conducted also show a degree of "discursive similarity" to the European approach unmatched by any of the other Nordic countries. Interviewees confirm the fact that the strategy has been created by "order" of the Commission and includes the areas that the Commission wants covered. The findings are perhaps surprising since Norway is not an official EU member but unsurprising for scholars of Norway's relationship to the EU: the Norwegian government has a long history of working closely with Brussels policymakers. This highlights how the European Commission has agenda-setting powers both within the EU and towards associated countries.

1. Introduction

ON THE 14TH of January 2020 at the Mesh coworking-space in Oslo, the Norwegian Minister of Local Government and Modernisation, Nikolai Astrup, presented *Norway's National Strategy for Artificial Intelligence* (NNSAI). The crowd consisted of 160 eager listeners from academia, civil society, trade and industry and the public sector. The tone of the presentation was optimistic: faced with the upcoming challenges of climate change, increasing globalisation and the age wave, Artificial Intelligence (AI) technologies are key in achieving smarter and more efficient use of resources so that Norway can maintain the same level of welfare and more efficiently deal with pressing societal challenges. The content of his presentation is reflected in the foreword of the NNSAI:

Artificial Intelligence represents vast opportunities for us as individuals, for business and industry, and for the public sector. If used optimally, technology can contribute to achieving the Sustainable Development Goals – not just in Norway, but globally.

Furthermore, Astrup recognised the fact that the United States and China were further along in creating consumer-oriented applications, although AI in Norway will differ in some key respects. He particularly emphasised Norway's own strengths in process industries: green shipping, aquaculture and petroleum activities as well as one of the top tier digitised public sectors in the world, Astrup states that by building on these country-specific strengths, Norway had the possibility to lead the way in AI technologies that are both human-friendly and trustworthy. The risks posed by AI were also mentioned in the foreword where some key risks are phrased as questions: accountability of AI systems, autonomous decision-making systems that potentially may cause harm and “how do we make sure that the technology does not intentionally or unintentionally perpetuate and reinforce discrimination and prejudice?”. The solutions for these potential AI threats lie in letting AI developments be guided by key principles such as transparency, explainability and cautious testing.²

Astrup is not alone in being an optimist about AI. In the last couple of years, we have seen an exponential increase in AI-related business investments, academic publications alongside an ever-increasing technical performance of AI technologies and systems.³ The perhaps most famous example of this comes from the chairman of the World Economic Forum (WEF), Klaus Schwab, depicting AI technology as key for the “fourth industrial revolution” akin to the role which the steam engine and electricity played for early industrialisation.⁴ As a new disruptive general-purpose technology, WEF predicts that AI will affect around 26% of jobs in emerging economies creating over 133

2 Ministry of Local Government and Modernisation (2020).

3 Perrault et al (2019).

4 Schwab (2017).

million new jobs globally by 2022.⁵ The Norwegian strategy should also be viewed in the light of these global developments. Increasing business investments, academic AI publications and technical performance has spurred a policy interest in AI whereby we can see an ever-growing increase in national AI strategies, policy documents, reports, and overall advocacy in the area. The OECD has set up an AI policy observatory as late as in early 2020 just to keep track of this development in which the organisation collects key data from AI policies around the globe. The OECD writes: “Artificial intelligence is at the top of policy agendas for governments and other stakeholder groups at both national and international levels”.⁶ This trend has not gone unnoticed by the Nordic countries. Figure 1 shows the mentioning of Artificial Intelligence in the Nordic governments official statements, policy documents, and press releases over the period 2010–2020.

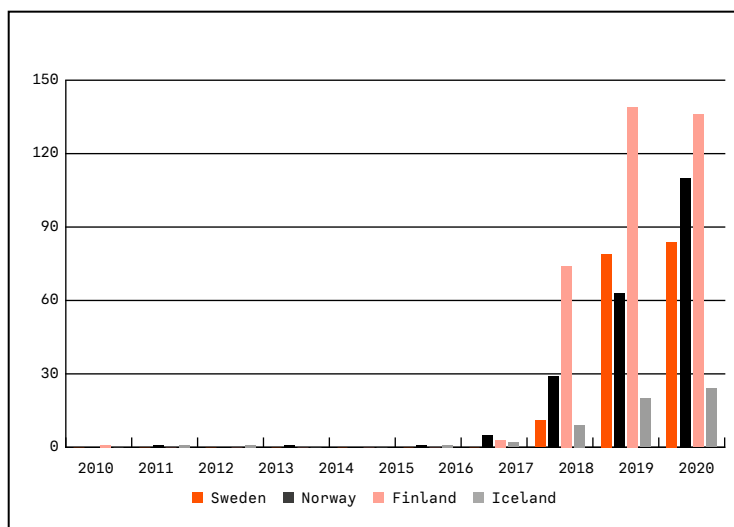


FIGURE 1: Nordic governments mention of “artificial intelligence” 2010–2020⁷

5 World Economic Forum (2019).

6 OECD (2020).

7 Denmark is unfortunately excluded in Figure 1 due to difficulties with the government’s online searching interface.

Furthermore, there are key concepts around AI visible in this policy agenda. As a main theme of this report, it is clear that *trustworthy AI* is one of these; a buzzword which permeates the policy sphere through official seminars, informal and formal discussions, and reports. Promoted by the European Commission (hereafter just “Commission”) and coined by the AI HLEG, set up by the Commission, trustworthy AI is at the core of the European coordinated approach on AI. Coordination, seen as a way to achieve a European approach to AI, is thought to be needed if the EU is to compete with the current global tech giants: The United States and China. The main question for this chapter is thus: to what extent is the Norwegian AI strategy coordinated with the European approach to AI? Unpacking this question not only gives content to the Norwegian approach, but also gives us insight into how and why the tightly coordinated European approach actually hopes to operate.

2. Norway and its neighbours

THE NORDIC COUNTRIES, in general, have been described as forerunners in both equality and environmental policy and innovation.⁸ The region also boasts a relatively lively and strong tech-sector which has gained traction of late and the Nordic AI market is expected to grow 35% annually over the coming years, surpassing 1 billion EUR in 2022.⁹

The Nordic integration towards the European project can be described as “differentiated integration”. The EU as we know it is increasingly diverse. EU membership cannot be viewed as a categorical variable. Rather, the EU increasingly resembles a multi-speed Europe whereby some countries are very tightly linked to the EU while other countries are not.¹⁰ In the Nordic countries, Iceland and Norway are not official members of the EU, while Sweden is an official member, Denmark is an official member bound by the financial pact, and finally Finland is officially within the Union as well as within the European Monetary Union (EMU). Together

8 Stende (2017).

9 Computer Weekly (2019).

10 Schimmelfennig & Winzen (2019).

with Iceland, Norway is thus the most loosely connected to the EU in the Nordics. At least formally. It is therefore a special case in many respects since it is not an official member of the European Union (EU) but still bears strong administrative ties to the Union both through policies and tight coupling of agencies and administration.¹¹ Both Iceland and Norway are members of the European Economic Area (EEA).

Thus, all the Nordic countries have a unique relationship with the EU. Despite the “Nordic block” commonly described as having similar democratic systems and societies – there are obvious differences when it comes to their relation to the EU. Moreover, the Nordic region and the Baltic states also have a long-standing institutionalised regional cooperation through the Nordic Council of Ministers (NCM) which also coordinated AI efforts in 2018.¹²

A litmus test for how closely the Nordics’ AI policies reflect the European approach to AI, Figure 2 shows the result of a word search for European AI language in Nordic AI policy documents. As the charts show, Norway’s AI strategy is the policy document with the highest number of references to the European approach, accumulating a total of 141 references through a word search query. Furthermore, the bar chart shows relatively high referencing to the word “trustworthy” – which, aforementioned, can be considered one of the core elements in the European approach to AI; this will be expanded upon in the next subsection. It is worth mentioning that the EU policy documents on AI were released in 2018 (European AI Strategy) and 2019 (AI HLEG), which partly explains why national policy documents prior to that date are less likely to use the language consistent with the EU approach.

11 Trondal & Kuhn (2018); Bauer & Trondal (2015); Eriksen & Fossum (2014).

12 Nordic Council of Ministers (2018).

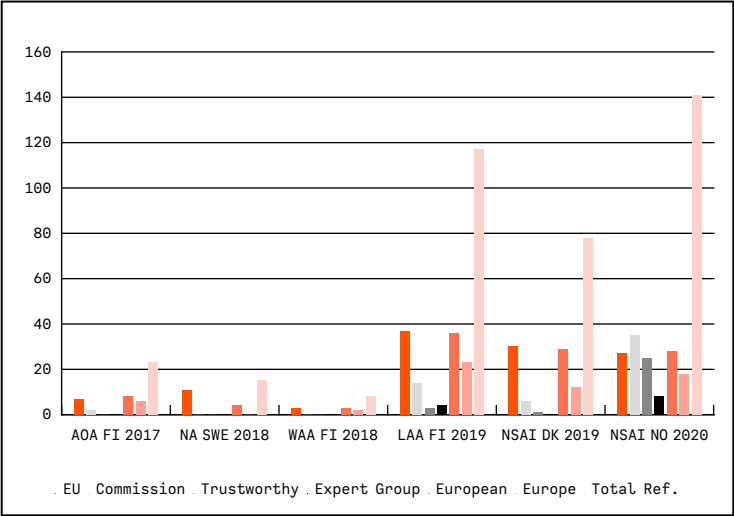


FIGURE 2: Nordic countries AI policy, references to a European approach¹³.

3. The National strategy for Artificial Intelligence

THE NORWEGIAN AI Strategy is a 67-page document which was published on January 14th 2020 and is thus the most recent of the Nordic strategies on AI (at the time when this text is written). The responsible ministry was the Ministry of Local Government and Modernization (KMD) and the writing group consisted of three full-time and one part-time government officials. In Norway, the AI policy was organised as a small writing group consisting of four people from the ministry,

13 The words are key concepts from the European approach such as trustworthiness and expert group as well as direct references to Europe the European Commission as well as the AI HLEG. The generic terms (commission, expert group, trustworthy) were manually controlled for false positives which then were weeded out from the search results. The abbreviations are the following: Finland's age of artificial Intelligence (AOA FI), Swedish National approach to artificial intelligence (NA SWE), Work in the age of Artificial Intelligence (WAA FI), Leading the way into the age of artificial intelligence (LAA FI), Danish National strategy for Artificial intelligence (NSAI DK), Norwegian National strategy for Artificial Intelligence (NSAI NO).

an interdepartmental working group tasked with providing feedback towards the writing group as well as a national feedback structure (inspillsmöten) consisting of relevant stakeholders from across all three sectors (public, private and civil society). The writing group was formed in late 2018 after the EU coordinated plan calling for member states to develop their own AI strategies, but also partly at the initiative of the then new Minister of Digitalisation (Nicolai Astrup). Interviews, conducted by the author, highlighted that the Minister of Digitalisation was very clear that AI strategy was a priority, which coincided with the visions of the government and the central administration. At the time, there was also a push from civil society stakeholders towards AI, who also emphasised that Sweden, Denmark, and Finland already had published their own AI strategies. The National Strategy contains five sections and 12 subsections, and the rest of this chapter is organised in line with its structure.

3.1. What is AI?

The Norwegian Strategy uses the definition of Artificial Intelligence promoted by the AI HLEG:

Artificial intelligence systems perform actions, physically or digitally, based on interpreting and processing structured or unstructured data, to achieve a given goal. Such systems can also adapt their behaviour by analysing and taking into account how their environment is affected by their previous actions.¹⁴

It not only borrows the definition but also the conceptualisation set up by the Commission with regards to the different AI technologies. Being a field of academic study, artificial intelligence incorporates machine learning, machine reasoning and robotics. The strategy furthermore makes a distinction between weak and strong AI also referred to as Artificial General Intelligence (AGI). In this conceptualisation, strong AI is considered to be able to do many differing tasks while weak or narrow

¹⁴ AI HLEG (2019b).

AI, due to its nature, is limited by its context. The strategy explains that the AI systems that existed at the time could interpret data from different devices, analyse those data and then perform actions. Moreover, the strategy explains machine learning and its subsections supervised, non-supervised and reinforcement learning. In such systems, the rules are deduced from the data on which the system is trained.¹⁵

3.2. A good basis for AI

This section is categorised under three subsections: regulation, data and language resources. The section is the largest of all the sections, and it tackles important questions around data and data sharing, regulation and how to fit AI into existing legislation or make way for AI. Data is recognised as forming the basis upon which AI is built, as access to high quality datasets are crucial for machine learning. The strategy makes a distinction between personal data which is covered by the Personal Data Act and open public data. The strategy reads:

No statutory obligation currently requires public sector data to be made accessible for use by others, but the goal is for data that can be made openly accessible to be shared so that it can be used by others (what we refer to as ‘reuse’).¹⁶

The strategy also introduces some methods for how data can be shared, including data lakes, data trusts, anonymisation interfaces, synthetic data as well as Application Programming Interfaces (APIs).

Building on earlier governmental reports,¹⁷ there are five specific sectors where the sharing and reuse of public data bears particular economic value: culture, research and education, government expenditure, transport and communications, and maps and property (geodata). As it is part of a bigger debate, the strategy states that the government plans to issue a White Paper on the data driven economy, and innovation.

¹⁵ Ministry of Local Government and Modernisation (2020).

¹⁶ Ibid., p. 14.

¹⁷ Ministry of Foreign Affairs (2019).

Furthermore, with regards to transparency, to “consider policy instruments that can make it easier for industry sectors to share data and that simultaneously safeguard privacy and data protection, security, and business interests”.¹⁸

The largest subsection is the one concerning regulation. The strategy points out that it has been a goal since its 2000 eRegulation project¹⁹ to regularly modernise existing legislation and make it technology-neutral so they can be applied when new technologies emerge. The dilemma concerning the “legal lag” is something that is pointed for example by Larsson.²⁰ However, this is just one of the issues within the regulatory realm. The strategy points out interoperability and “personal data: consent and statutory authority” as two areas which present specific regulatory challenges.²¹ It is worth noting that the strategy includes starting up more regulatory sandboxes to test new technologies and/or business models within specific parameters.²² For instance, the Norwegian maritime authorities started their first testbed for autonomous vessels in 2016 and since then two more have been approved.²³

Thus, the need for up-to-date regulation, data sharing capabilities as well as building AI on existing Norwegian languages including Sami, is well recognised. Acknowledging that different forms of language processing are some of the areas most central in AI technologies, the strategy highlights the need for language resources available to developing “homegrown” AI technologies in this field.

3.3. Developing and Leveraging AI

Globally, we have seen increased investments in AI related technologies, start-ups and research.²⁴ Perhaps the most famous depiction

18 Ministry of Local Government and Modernisation (2020), p. 18.

19 Norwegian Ministry of Trade and Industry (2000-2001).

20 Larsson (2020).

21 Ministry of Local Government and Modernisation (2020), p. 21.

22 Ministry of Local Government and Modernisation (2020), p. 24.

23 Norwegian Maritime Authority (2017).

24 Perrault, et.al (2019).

of the transformative potential of AI, is by the chairman of the World Economic Forum, Klaus Schwab, who describes AI as a key technology in “the fourth industrial revolution”.²⁵ One key focus in both the NNSAI and in the EU approach is the potential that AI has for the future. This focus is visible in their discussions on how to harness the potential of AI related technologies, funding development, knowledge and implementation of AI. There are seven uses of the word “funding” in the Norwegian strategy of which five are under the chapter 3 “Developing and leveraging AI”. The chapter is split in two sections, concerning a) Research and Higher Education and b) Skills.

3.3.1 Research and Higher Education

In the Norwegian strategy – in accordance with the European approach – there is focus on investment in AI research. The perhaps most speaking graph for these developments is a chart that shows the expenditure from the Norwegian Research Council (NRC). As suggested by Figure 3, the overall trends point towards an increase in research and development in AI in Norway which also is in line with the EU which increased its AI related investments with 49% from 2018 to 2019 and reached 5.2 billion USD.²⁶

²⁵ Schwab (2017).

²⁶ International Data Corporation (2019).

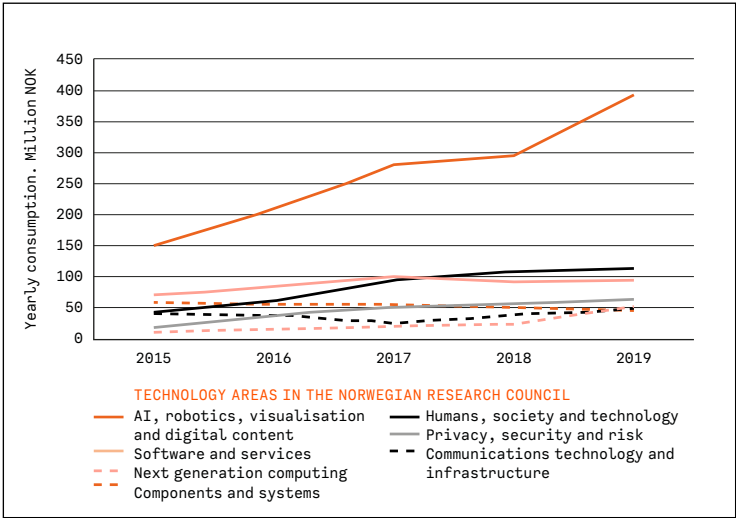


FIGURE 3: Yearly expenditure of Norwegian Research Council.

3.3.2 Skills

The NNSAI recognises that the development of technological skills is vital to achieve a good basis for AI in Norway. Of the many changes underway, the most vital one is the Norwegian Curriculum Renewal, whereby natural sciences and mathematics will include programming and computational thinking skills. Furthermore, the existing further education programmes will be reinforced with a new Skills Programme in 2020. The NNSAI recognises the importance of the existing workforce to learn and adapt, hence why the Norwegian Government presented a skills reform for lifelong learning in April.

3.4. Enhancing innovation capacity using AI

The *European Innovation Scoreboard 2020*²⁷ shows that the Nordics rank the highest and most innovative of all the EU countries. Numbers from the OECD published in January 2020 show that Norway is in third place out of all OECD countries with regards to the most innovative firms.²⁸ The NNAIS suggests that Norway can build on its existing strengths,

²⁷ European Commission (2020).

²⁸ OECD, (2020).

such as; health, oil and gas, energy, the maritime and marine industries and the public sector, and advance these through the use of AI. Public sector agencies are already exploring the potential in AI related technologies for example in the Norwegian tax administration,²⁹ as well as the Norwegian State Educational Loan Fund.³⁰ Since it is acknowledged that increased co-creation between the public and the private sector is vital to unlock the innovative potential of applying AI, the Norwegian government has therefore committed to establishing a new programme for interaction between start-ups and the public sector.³¹

The NNSAI furthermore follows the policy instrument Digital Innovation Hubs (DIHs), suggested by the Commission in 2016, to get Small and Medium sized enterprises on board with implementing AI technologies. Since then, four DIHs have been initiated in Norway. The NNSAI recognises that a new generation of larger and more binding DIHs will be created. Another key method is to find new and innovative forms of collaboration between public and businesses, hence why the NNSAI also suggests that the government should engage in a dialogue with DigitalNorway – a non-profit organisation working to harmonise digital transformation within Norwegian businesses.³²

3.5. Trustworthy AI

The European priority for trustworthiness is visible in the NNSAI, notably in the foreword by the then Minister of Digitalization, Nikolai Astrup:

Norway enjoys a high level of trust and some fundamental values that permeate our society. We respect human rights and privacy, [...] This is something we perhaps take for granted in Norway, but leading the way in developing human-friendly and trustworthy artificial intelligence may prove a vital competitive advantage in today's global competition.³³

29 OECD (2016).

30 Norwegian State Educational Loan Fund (2017).

31 Ministry of Local Government and Modernisation (2020).

32 Ibid.

33 Ibid., foreword.

An entire chapter in the NNSAI is dedicated to the notion of trustworthiness. The chapter begins by describing four specific issues with AI: “big data vs data minimisation”, “data quality”, “lack of transparency” and “autonomy”. The themes are very upfront: AI needs vast amounts of data for training but at the same time, the principles for data protection is data minimisation. On the one hand, machine learning algorithms need vast amounts of data for the machine to be able to learn. On the other hand, contemporary regulation such as the Personal Data Act in Norway and the GDPR at the European level is focused on minimising access to and usage of certain types of data. The NNSAI acknowledges this as a dilemma: “consequently, the need for large datasets can conflict with the principle of data minimisation”. Although there are also ways around this problem through, for example, encryption as well as anonymised data or synthetic datasets.³⁴

The quality of the data is of great importance for training a machine learning system since errors in the data can have effects on the analysis performed. The lack of transparency is due to the fact that some deep learning algorithms can be considered as “black boxes” since even their creators do not know why they make certain decisions and what information they base their decisions on.³⁵ In this light, autonomy is obviously a possible challenge since it raises the question of the accountability of decision-making algorithms.³⁶

The fact that “Trustworthy AI” – which is at the centre of the European approach – is mentioned in the beginning as well as granted an entire chapter in the strategy demonstrates the fact that the Norwegian approach is tightly linked to the European approach. The notion of Trustworthy AI permeates the strategy which contains the word 25 times. Whereas the rest of the Nordic AI policy documents contain the word four times in total. Furthermore, the strategy clearly adopts the seven principles put forth by the AI HLEG for the realization of trustworthy AI.³⁷

³⁴ Ibid.

³⁵ Larsson & Heintz (2020).

³⁶ Ministry of Local Government and Modernisation (2020), pp. 57-58.

³⁷ AI HLEG (2019b).

4. Summary and discussion

THE NORWEGIAN STRATEGY is a comprehensive, well-written document based on both its country specific strengths and to the existing coordinated European framework for AI. The strategy covers the areas that are the most pressing as well as most common aspects of AI. The EU influence is apparent, as one interviewee put it:

It has been proven that our strategy covers the areas where the Commission has called for. It is a kind of order to the Member States in the coordinated plan and that the strategy created should cover specific areas. Basically, they cover those areas.

As stated in the introduction, although Norway is not an official member of the EU, the administrative side of the government harmonised with EU policy frameworks. This is also true when it comes to the NNSAI. The late strategy, published in 2020, and extensive referencing to the EU approach suggest that the Norwegian strategy is intended to follow the European approach. This was confirmed by interviews in which it was emphasised that the strategy is essentially a document in line with what the Commission encouraged in the coordinated plan back in 2018. Compared to other high-level AI strategies/policies in the Nordic states, Norway is one of the most tightly discursively integrated to the European approach on AI as well as the national strategy most guided by the AI HLEG. We can conclude that the Norwegian AI strategy has a high degree of similarity to the European approach generally and to the documents produced by the AI HLEG specifically. The AI HLEG has been an influential actor in this regard which also was pointed out by interviewees who expressly emphasised on the Ethical Guidelines for Trustworthy AI³⁸ and, to a lesser extent, the Policy and Investment recommendations.³⁹

A part of establishing national strategies from the EU level, States are able to draw on the EU's strong geopolitical position vis-à-vis the US

³⁸ Ibid.

³⁹ AI HLEG (2019a).

and China and thus to be able to leverage the EU in the global “AI race”. This is apparent in the Coordinated plan from 2018 which states: “Amid fierce global competition, a solid European framework is needed”.⁴⁰ As part of this endeavour, member states were encouraged to find examples of how they implement AI and, most importantly, member states were encouraged to “dig up” quantifiable measures to be able to show how much they invested in AI. One interviewee portrayed this as an exercise in “creative accounting”. It was described as a struggle from the policymakers in Norway, faced with the dilemma of trying to track down numbers on how much research is put into AI and how many PhDs and master students write about AI:

The Commission has also been busy positioning the EU vis-à-vis China and the United States and to make visible a major investment in Europe. They have been very much looking for “numbers” – quantification of research, investments, study places, candidates.

This raises two distinct questions towards the nature of this and also other AI strategies. Will the member states (Norway in this instance) follow up on what they have set out to do in the strategy? And to what extent do these strategies serve to legitimise the EU’s geopolitical objectives?

It has been proclaimed by many, including in the NNSAI, that allegedly (which also has been emphasized earlier in this chapter) the potential of AI lies in the future. Here, it is too early to tell whether the policy approach will foster a robust and strong AI sector in Norway. Although the NNSAI is grand in its ambitions, it could be argued that Norway has not made as much progress as some of its Nordic neighbours: Finland, which has been driving the European approach to an extent through the chairman of the AI HLEG (Pekka Ala Pietilä), also has produced three reports on AI since 2017 and has involved a large network of corporate interested parties into the policy process. In Norway, stakeholders from all three sectors (academia, civil society

⁴⁰ European Commission (2018a).

and corporate/business) were consulted but not as deeply integrated into the policy process as in Finland.

Within the organisation theory,⁴¹ separating the talk and the walk is common under the concept “decoupling”.⁴² This theory suggests that how organisations describe themselves can be decoupled from organisational practice and that organisational capacity stands in direct relation to efficient implementation. Yet, although we can already see some areas in Norway which are “up and running” and the strategy shows clearly how the government needs to, and plans to provide, a good foundation for developing trustworthy AI, it remains to be seen how this will play out in practice.

41 March & Olsen (1989).

42 Meyer & Rowan (1977); Brunsson & Adler (1989).

References

- AI HLEG (2019a) Independent High-Level Expert Group On Artificial Intelligence, *Policy And Investment Recommendations For Trustworthy AI*, Available at: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- AI HLEG (2019b), Independent High-Level Expert Group On Artificial Intelligence Set Up By The European Commission, *Ethics Guidelines For Trustworthy AI*, Available at: <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>
- Bauer, M., & J. Trondal, J. (red.) (2015). The Palgrave handbook of the European administrative system. Cheltenham: Palgrave Macmillan.
- Brunsson, N. & Adler, N. (1989). The Organization of hypocrisy: Talk, decisions, and actions in organization. Chichester: Wiley.
- Computer Weekly (2019). *Venture capitalists driving Nordic tech enterprise growth*. Available at: <https://www.computerweekly.com/news/252459375/Venture-capitalists-driving-Nordic-tech-enterprise-growth>
- Eriksen, E. O. & Fossum, J. E. (red.) (2014). Det norske paradoks. Oslo: Gyldendal Akademisk.
- European Commission (2018a). European Commission Brussels, 25.4.2018 Com (2018) 237 Final, Communication From The Commission To The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions, *Artificial Intelligence for Europe* {SWD(2018) 137 final} Available at: <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>
- European Commission (2018b). European Commission Brussels, 7.12.2018 Com(2018) 795 Final, Annex To The Communication From The Commission To The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions Coordinated Plan On *Artificial Intelligence Coordinated Plan on Artificial Intelligence*, Available at: <https://ec.europa.eu/digital-single-market/en/news/coordinated-plan-artificial-intelligence>
- International Data Corporation (2019). Automation and Customer Experience Needs Will Drive AI Investment to \$5 billion by 2019 Across European Industries. Available at: <https://www.idc.com/getdoc.jsp?containerId=prEMEA44978619>.
- International Data Corporation (2019) Nordic Digital Insights - IT Market Predictions 2019. Available at: <https://www.nordic.idc.com/promos/RESOURCES/ATTACHMENTS/Nordic-Digital-Insights-Report.pdf>
- Larsson, S. (2020). On the Governance of Artificial Intelligence through Ethics Guidelines. *Asian Journal of Law and Society*, 1-23.
- Larsson, S., & Heintz, F. (2020). Transparency in Artificial Intelligence. *Internet Policy Review*, 9(2).
- March, J. G., & Olsen, J. P. (1989). Rediscovering institutions: The organizational basis of politics. Enskede: TPB.
- Meyer, J., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), 340-363. Retrieved September 1, 2020, from <http://www.jstor.org/stable/2778293>

Ministry of Foreign Affairs (2019). Report to the Storting no. 27 (Meld. St. 27 (2018–2019)). Available at: <https://www.regjeringen.no/en/dokumenter/meld.-st.-27-20182019/id264250/>

Ministry of Local Government and Modernisation (2020) National Strategy for Artificial Intelligence. Norway. [Kommunal- og moderniseringsdepartementet (2020) Nasjonal strategi for kunstig intelligens].

Nordic Council of Ministers (2018). *AI in the Nordic-Baltic region*.

Norwegian Ministry of Trade and Industry Ot.prp. nr. 108. (2000-2001). Draft resolution and Bill to amend various acts in order to remove obstacles for electronic communication.

Norwegian State Educational Loan Fund (2017). Final report from the concept study of the rationalisation potential of AI.

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., and Niebles, J. C. (2019). "The AI Index 2019 Annual Report", AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA. Available at: <https://creativecommons.org/licenses/by-nd/4.0/legalcode>

Organization for Economic Co-operation and Development (OECD) (2017). Digital Government Review of Norway BOOSTING THE DIGITAL TRANSFORMATION OF THE PUBLIC SECTOR.

Organization for Economic Co-operation and Development (OECD) (2016). Advanced Analytics for Better Tax Administration: Putting Data to Work, OECD Publishing, Paris. Available at: <https://doi.org/10.1787/9789264256453-en>

Organization for Economic Co-operation and Development (OECD) (2020). Business innovation statistics and indicators. Available at: <https://www.oecd.org/innovation/inno/inno-stats.htm#indicators>

Organization for Economic Co-operation and Development (OECD) (2020). OECD Policy observatory website. Available at: <https://oecd.ai/countries-and-initiatives>

Schimmelfennig, F. & Winzen, T. (2019). Grand theories, differentiated integration, Journal of European Public Policy, 26:8, 1172-1192, DOI: 10.1080/13501763.2019.1576761

Schwab, K. (2017). The fourth industrial revolution. London Portfolio Penguin.

Sjøfartsdirektoratet. (2017). Horten blir testområde for autonome skip [Norwegian Maritime Authority (2017): Horten to be test bed for autonomous ships]. Available at: www.sdir.no/en/

Stende, T. (2017). Is the Nordic Region best in the world? Copenhagen. Available at: <https://doi.org/10.6027/ANP2017-776>.

Trondal, J. & Kühn, N. (2018). Norsk Europaforvaltning – Et organisasjonsteoretisk perspektiv på departementenes EU-arbeid; Norsk statsvitenskapelig tidsskrift 02-03 / 2018 (Volum 34).

World Economic Forum. (2018). The Future of Jobs Report, Centre for the New Economy and Society, Available at: http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

AI policy in the Nordics

Pledging openness, transparency
and trust, while expressing
readiness to apply AI in society

Stephen Cory Robinson

Cory Robinson, Ph.D., is a Senior Lecturer/Assistant Professor of Communication Design at Linköping University in Sweden. Focusing on areas of privacy and self-disclosure, his research investigates disclosure of personal data in ecommerce, and ethical frameworks for collection of personal data in wearables and IoT. Cory is also interested in how culture affects perceptions of privacy and what is considered private information.

Summary

THIS CHAPTER PRESENTS a condensed version of selected findings from a recent publication.¹ By examining four national strategic plans for developing AI, the author explores the distinctive differences in how Nordic countries (Denmark, Finland, Norway, Sweden) position themselves using their unique cultural values (i.e., openness, transparency and trust) as business principles to support development and deploying AI technologies throughout their societies. As an example of enshrining cultural values in policy, the value of trust requires purposeful action and increases over time through several interacting societal processes – but can also be lost quickly, especially in scenarios where digital trust

1 Robinson (2020).

has been eroded (i.e., hacking, fraud, or technological incompetence by industry or government institutions). The sharing of values of openness, transparency and trust by all Nordic countries creates a cultural unity amongst them, affording an opportunity to demonstrate how shared cultural values can encourage strong, value-laden technology policy. This chapter explores the role these three cultural values play in developing national strategies for implementing and using AI. The policy review reveals Norway and Finland assert the value of public AI education, while Denmark's policy clearly illustrates AI case studies in Danish society – however, the brevity, generality, and vagueness of the Swedish policy guide is concerning. The analysis reveals that openness, transparency and trust do influence Nordic AI policies, while themes of democracy, ethics and privacy are also prominent in the policies.

1. Introduction

DENMARK, FINLAND, NORWAY, and Sweden are unique countries, and they all share much in common: all are EU member states or part of the EEA, and based on analytical cultural comparison theories, such as those laid out by Geert Hofstede, these countries share relatively similar cultures.² Each country is also a member of the Nordic Council, the world's oldest regional intergovernmental partnership.³ Additionally, they share a unique bond through cultural identity,⁴ high levels of openness in society,⁵ transparency with regards to information access,⁶ and trust.⁷

This chapter presents a condensed version of selected findings from a recent publication,⁸ and examines the national strategic plans for AI development of Denmark, Finland, Norway and Sweden in order to explore the distinctive ways in which the Nordic countries use their

2 Hofstede (2019).

3 Robinson (2020).

4 Smith et al. (2003).

5 Vesa (2015).

6 Jorgensen (2014).

7 Ortiz-Ospina & Roser (2020).

8 Robinson (2020).

unique cultural values – trust, transparency, openness – as business principles to support the development and deployment of AI technologies in their societies. After conceptualising cultural values, and then exploring the unique cultural environment of the Nordic countries and their cultural values, the chapter charts how the values of openness, transparency and trust are present in Nordic AI policy documents. This chapter concludes with a discussion of the Nordic policy documents and some implications for public policy.

1.1 Shared cultural traits in the Nordics

The aforementioned four Nordic countries have high levels of openness. In defining a set of personality traits, being open means an openness to experience, demonstrated by one's willingness to engage in new activities or ideas, and being naturally curious.⁹ Openness can be framed in the context of policies focusing on citizen engagement and citizens' access to information, or in the context of policies emphasising access on equal terms including the lowest cost, or no more than cost of dissemination, with access being user-friendly, opportune and timely.¹⁰ Societies that lack regard for openness might abuse power, while societies that value openness enhance civic cohesion and overall system performance.¹¹ Being responsive to innovative ways of thinking is an attribute of open governments, where demands from citizens and other stakeholders are acted upon and the government is accessible at all times.¹² Another definition of openness applicable to AI development can therefore encompass "accessibility of knowledge, technology and other resources; the transparency of action; the permeability of organisational structures; and the inclusiveness of participation".¹³

Transparency, another shared cultural trait in the Nordics, is crucial for creating trustworthy governments, but it is an element often

9 Robinson (2018).

10 OECD (2007) & OECD (2019).

11 Gotz (2015).

12 OECD (2019).

13 Schlagwein et al. (2017).

overlooked when exploring differences between cultures.¹⁴ The value of transparency can be defined as citizens' ability and right to access information about and produced by their government.¹⁵ Nordic nations practice high levels of transparency,¹⁶ and they are highly regarded within the EU for their innovations in guaranteeing the transparent operation of government.¹⁷ More specifically, transparency involves the extent to which an organisation or individual discloses relevant information about their performance, functioning, decision-making processes and procedures.¹⁸ The availability of information about the performance of a business or the internal workings of an entity are some of the many components that transparency might include, and they enhance the abilities of individuals or external organisations to monitor activities or decisions taking place within a given organisation.¹⁹ A holistic definition of transparency is "the availability of information about an organisation or actor that allows external actors to monitor the internal workings or performance of that organisation."²⁰

Lastly, the concept of trust involves potentially exposing one to a vulnerability of some sort,²¹ and it functions at various levels of society, being interpersonal, or relating to individuals' trust in organisations. The latter has been discussed by some as institutional trust,²² such as trust in government institutions²³ or public institutions.²⁴ World rankings of "happiness" indices, such as the World Happiness Report, conceptualise trust as social trust, or the notion that people in general can be trusted.²⁵

14 Grimmelikhuijsen et al. (2013).

15 Hood (2006).

16 Jorgensen (2014); Transparency International (2015).

17 Globaliseringsrådet (2008) & Bunyan (1999).

18 Curtin & Meijer (2006) & Welch, Hinnant & Moon (2005).

19 Grimmelikhuijsen et al. (2013).

20 Ibid.

21 Abbass et al. (2016).

22 Sonderskov & Dinesen (2016).

23 Torney-Purta, Henry Barber & Richardson (2004) & Stoyan (2016).

24 Stevenson & Wolfers (2011) & Marozzi (2015).

25 Berggren & Bjørnskov (2011).

2. Findings and analysis country-by-country

TO EXPLORE THE inclusion of openness, transparency and trust in the Nordic policy documents, I have applied textual analysis, a systematic procedure for analysing documents that explores questions of how texts reflect, reject, or influence societal views.²⁶ In short, the researcher interprets texts by connecting textual messages to larger societal elements that suggest the meanings that might be gleaned by different individuals.²⁷ For this textual analysis, the primary texts, or items of main focus,²⁸ are the Nordic national AI policy documents, while the secondary texts are related journal articles and other national technology policy documents serving to support or challenge information presented in the primary texts. In the first cursory examination, I ran keyword searches, using quantitative keyword search components to lead to relevant portions of the texts for deeper qualitative analysis. Partial word strategies were implemented to account for variances in context, word tense and plurality. In the cursory and thorough examination, the central themes of openness, transparency and trust were colour coded for visual clarity, and efficient referencing when inspecting the documents.

Each of the four national strategic guidelines for AI incorporates, to some degree, the cultural values of openness, transparency and trust. This section analyses how those values are included in each of the four documents. The following sections will analyse each document, including *how* these cultural values were or were not upheld in national AI strategies.

²⁶ Hawkins (2017).

²⁷ Ibid.

²⁸ Bainbridge, Goc & Tynan (2008).

Table 1: Total references in four national AI strategy documents				
Search value	Denmark	Finland	Norway	Sweden
Openness	1	1	0	0
Transparen*	10	15	14	3
Trust*	6	41	40	2

* Partial word strategy to account for variance in context, word tense, or plurality

2.1 Analysis of the Danish national AI policy

The Danish policy is only one of two documents that explicitly states openness, in the section “Explainability” of AI it asserts: “however, the public authorities have a special responsibility to ensure openness and transparency in the use of algorithms.”²⁹ Though transparency is mentioned, the same paragraph states that “explainability is not the same as full transparency of algorithms, as there are business interests in the private sector, for example.”³⁰

Highlighting notions of transparency, but not as thorough as Norway’s explanation of AI and related technologies, Denmark’s document does provide some technical definitions and explanations such as: “What is artificial intelligence?”³¹ “What are cloud technologies?”³² and, sections detailing issues of “Challenges for the use of artificial intelligence in Denmark”³³ or applications of AI, for example “Case: Artificial intelligence provides better food safety.”³⁴

The notions of trust are undeniably highly regarded in the Danish culture:

29 Denmark’s Ministry of Finance and Ministry of Industry, Business and Financial Affairs (2019), p.28.

30 Ibid.

31 Ibid., p. 6.

32 Ibid., p. 39.

33 Ibid., p. 16.

34 Ibid., p. 46

The Danish population has a high degree of trust in each other, and we are generally positive towards digital and technological development.³⁵

Denmark's national strategy for AI references 'trust' five times, with one mention of 'trustworthy', which specifically assures that the government will keep AI, its algorithms and results trustworthy.³⁶ Here, it is unique that assurances of trustworthy (1) AI and (2) algorithms are mentioned separately, and not under the typical umbrella term of AI (which can, incorrectly, lump machine learning, algorithms, automation and object recognition into one concept). However, ensuring AI and its algorithms are trustworthy is notable, as it would be "easier" to simply state "the Danish government will ensure trustworthy AI" simplifying the technologies and responsibilities. But there are stark differences in how and why the Danish government should (and must) ensure trustworthy AI and, separately, the algorithms. Unfortunately, why did the policy exclude trustworthy machine learning or automation?

Similar to many states, Denmark has suffered its fair share of public debacles. Gladsaxe is a Danish municipality that became infamous in late 2018, as its civil servants tried to pool data from various public registers, and implement profiling algorithms and machine learning in order to identify children at risk of abuse and/or neglect. The aim was to identify families with children, and to create an algorithm that could assign risk probability scores. The project was, however, put on hold after these ambitions were made public.³⁷ Even in the face of several questionable incidents like these, and attendant policies,³⁸ the Danish government has still been willing to implement innovative and bold policies. Danish policy events should serve as an example for other governments of how poorly implemented and communicated policies can quickly erode public trust. Assuring trustworthy AI and trustworthy algorithms indicates the government's proactivity

35 Ibid., p. 7

36 Ibid., p. 6

37 Sorgenfri Kjær (2019).

38 Alfter (2019) & Byrne & Sommer (2019).

in ensuring implementation of AI policy in Danish society does not erode high levels of trust in government.

2.2 Analysis of the Finnish national AI policy

The Finnish policy document is similar to the Danish policy document, in that it also explicitly states openness, but the term is mentioned only once in the entirety of the document. Moreover, this mention of openness is in context to the Finnish economy,³⁹ not as a cultural value.

Similar to other documents in the analysis, mentions of democracy are present; Finland's policy has an entire section devoted to AI and democracy, however, mentions of democracy are notably missing from this entire two-page section. It is not until one page later that democracy is plainly mentioned,⁴⁰ but only within a five bullet-point list summarising the recommendations from the Chapter 2, "International AI experts: Towards the third wave of artificial intelligence". The single bullet point referencing democracy, states: "Respect the principles of democracy and freedom",⁴¹ summarising how democracy is important in "Western" nations, but balancing this with stakeholder's "benefits" derived from AI.

The Finnish policy document situates democracy and the benefits of AI as something that can be balanced. "Solutions based on artificial intelligence should be seen as a way of reinventing society and increasing citizens' participation in decision-making and democratic processes"⁴² – it references democracy as being a component in the process, but does not provide clarity as to how it will be upheld. In Chapter 3 of the document, "Eleven key actions ushering Finland into the age of artificial intelligence", democracy is highlighted in how democracy (along with environmental affairs) can be promoted if investments in AI were made,⁴³ how concerns of AI raises issues affecting human rights and

39 Ministry of Economic Affairs and Employment of Finland (2017), p. 76.

40 Ibid., p. 38.

41 Ibid., pp. 38-39.

42 Ibid., p. 39.

43 Ibid., p. 80.

democracy,⁴⁴ or ethical issues of democracy inherent in AI systems and automated decisions.⁴⁵ While it is stated that “AI ethics must not be seen as a factor posing limitations on the activities only, but also as a factor that creates something new, and provides increasing opportunities,”⁴⁶ the aforementioned references invoke a sentiment that democracy is a component limiting the potential of AI.

Notions of transparency are mentioned numerous times in the document, the most of the four explored policy documents (though, Finland’s is the longest, being 62 pages longer than the second longest, Denmark’s). Two mentions are within case studies of Finnish companies implementing AI technologies,⁴⁷ with the first conceptual mentions relevant analysis in the section, “Artificial intelligence as a factor renewing society and Democracy”⁴⁸ where transparency of decision making in AI algorithms are described as concerns in Finland and elsewhere.⁴⁹

Next, transparency is described in the context of the black box, where algorithms should be transparent processes. Importantly, the document tries to downplay the relevance of this concern, “But is this [algorithmic transparency] really needed and is it even possible?”⁵⁰ through use of technological analogy. By noting how humans do not necessarily understand how 4G technology handles transmitting of calls across different continents, the document (controversially) argues transparency of processes is not relevant, because “we trust the data communications system and the parties operating it.”⁵¹ “Here too, it is a question of trust rather than transparency”⁵² – conversely, by devaluing and downplaying citizens’ legitimate concerns, one also potentially erodes citizens’ trust.

44 Ibid., pp. 103 & 106.

45 Ibid., pp. 103 & 123.

46 Ibid., p. 106.

47 Ibid., pp. 19 & 21.

48 Ibid., p. 35.

49 Ibid., p. 108.

50 Ibid., p. 36.

51 Ibid.

52 Ibid.

The importance of trust and its place in Finnish society is clear: No other European country ranks higher in citizen trust, with 85% of respondents agreeing to the statement “generally speaking, most people in their country can be trusted.”⁵³ However, this unique characteristic of Finnish society cannot be taken for granted:

It practically obliges us to an active approach, understanding of the prerequisites of trust in the age of artificial intelligence, and agile implementation both nationally and as part of the international community.⁵⁴

Notions of trust in the Finland policy document are quite diverse, ranging from trust in the healthcare sector with adoption of AI, notion of trust as a factor in democracy, and its importance in a future Finland with deep integration of AI. The document notes that creating a trust-based society and maintaining trust are indeed separate actions, both entailing participation by citizens and government. On a similar note, but separate section of the document, “Artificial intelligence as a factor renewing society and democracy”, trust is critical: “Do we trust the organisations and people producing the AI-based services? Do we trust the technology solutions offered to us?”⁵⁵

The Finnish AI strategy addresses the crucial balance of supporting and maintaining societal trust while implementing AI technology in government services. By mentioning trust and how AI might affect this in futuristic Finland, it creates even more pressure for trust to be implemented and supported by AI by noting how the futuristic society must have (1) trust in their own society, (2) maintain trust between citizens and (3) how Nordic countries must remain trust-based societies. By presenting this “forward looking” view, the policy document seems to emphasise the importance of trust in society, and possible pitfalls with AI in society, clearly underscoring how a technology like AI can

53 Eurobarometer (2018).

54 Ministry of Economic Affairs and Employment of Finland (2017), p. 103.

55 Ibid., p. 36.

erode trust between people (social trust) and in one's society (institutional trust). Finland, like the Norwegian policy document, highlights the value of public AI education – mentioning the “Elements of AI” Internet-course in the context of “Versatile education programs.”⁵⁶ The course should be seen as a framework for other nations to educate citizens, as more than 100,000 Finns completed the course within months of its launch.

2.3 Analysis of the Norwegian national AI policy

An open government is responsive to innovative ways of thinking and demands from citizens and other stakeholders, and is accessible at all times, to all individuals.⁵⁷ A pattern repeats itself here, similar to Sweden's document, as there are no explicit mentions of openness. However, there are direct mentions to democratising AI and its processes: “[AI] that is developed and used in Norway should be built on... democracy”.⁵⁸ ‘Democracy’ is mentioned a total of five times in the document, including AI must foster a democratic society,⁵⁹ and AI must have no adverse consequences for democracy.⁶⁰ Even in the absence of openness, it is clear that democracy and openness of AI is a necessity for deploying AI in the Norwegian society and government.

By explaining in four pages of the policy strategy what AI technologies entail, a message of clarity and transparency about AI technology and Norway's proposed use of the technologies is presented. Indeed, the document details the policy document as “intended for the civilian sector – both private and public”.⁶¹ Crucially, the policy document informs of the “black box” problem,⁶² and identifying “Lack of transparency” as an issue that might be resolved in two ways: (1) not all systems are “black boxes” or in systems where explainability is important,

56 Ibid., p. 75.

57 OECD (2019).

58 Norway's Ministry of Local Government and Modernisation (2020). For an analysis entirely dedicated to the Norwegian AI strategy, see Frans af Malmberg's chapter in this volume.

59 Ibid., p. 59.

60 Ibid., p. 60.

61 Ibid., p. 8.

62 Ibid., p. 12.

deep learning might be more appropriate,⁶³ and that (2) explainable AI, or encouraging the explainability of black box algorithms, can analyse data's significance for an outcome, or what significance other elements might have (resulting in clear logic behind the outcome).⁶⁴

Trust is primarily discussed within five sections of the document, including “Data and data management”, “Regulations”, “Research and higher education”, “Industrial policy instruments”, and “Security”. Notably, the Norwegian policy is an outlier, as trust has its own 8-page chapter in the policy document, titled “Trustworthy AI”. The theme of ‘trust’, mentioned 40 times, is seemingly important, and its inclusion and importance within AI frameworks is clear.

When speaking of automated decision making in providing citizen services, the document states:

More consistent implementation of obligations can lead to higher levels of compliance and to a perception among citizens that most people contribute their share, which in turn can help build trust.⁶⁵

In the chapter “Trustworthy AI”, the policy document states “research, development and use of artificial intelligence in Norway should promote responsible and trustworthy AI” and “supervisory authorities should oversee that AI systems in their areas of supervision are operated in accordance with the principles for responsible and trustworthy use of AI”.⁶⁶ Trustworthy AI is a prominent theme in the document, and if implemented appropriately, might reinforce notions of social trust and trust in organisations and government: “The [Norwegian] Government wants to maintain and strengthen this [social and organisational] trust at the same time as artificial intelligence is adopted in new and innovative ways”.⁶⁷

63 Ibid., p. 58.

64 Ibid.

65 Ibid., p. 27.

66 Ministry of Local Government and Modernisation (2020), p. 56.

67 Ibid.

In several aspects of its national AI policy document, Norway is indeed a vanguard – for example, no other Nordic policy mentions data trusts or synthetic data. In terms of openness of data, Norway has provided clear, applicable examples of how sharing data is possible between different stakeholders within national policy for implementing AI. The policy document provides a foundation for any stakeholder to possess basic AI literacy and therefore process and rationalise the proposed strategies in the document. Norway should be commended here, as the other Nordic policy documents either (1) do not explain AI technologies, or (2) only provide a concise or highly technical overview of AI technologies. The explanation of trustworthy AI and how the government will approach implementing and supporting trustworthy AI is well reasoned and should be commended. Versus their southern neighbour, Denmark, one might reflect Norway's implementation of AI policy, through clear communication and how they will enforce trustworthy AI, contrasts positively with Denmark's mixed record on AI policy.⁶⁸ Norway's policy document excels in explaining what trustworthy AI is, and how the government will approach implementing and supporting trustworthy AI. Norway's stance towards public AI education is positive, asserting the government will make the Norwegian-Finnish developed online course, "Elements of AI", available in the Norwegian language from 2020.

2.4 Analysis of the Swedish national AI policy

As stated by the OECD, an open government is responsive to innovative ways of thinking and demands from citizens and is accessible at all times, to all individuals.⁶⁹ However, how does one establish policies building on this principle when openness is not explicitly stated in Sweden's national AI strategy document? Overlooking this exclusion, the policy document does have statements that indirectly relate to openness. For example, the challenges associated with AI (i.e., discrimination, loss of trust, financial damage and lesser functioning democracy) are referenced, but solutions are not offered: "The

⁶⁸ Alfter (2019) & Byrne & Sommer (2019).

⁶⁹ OECD (2019).

countries that succeed in harnessing and realising the benefits of AI while managing the risks in a responsible manner will have a great competitive advantage internationally”.⁷⁰ Potentially negative consequences for democratic processes, and relatedly openness, are important to address. Sweden importantly notes disinformation as a realistic threat to democracy, and openness of society: “AI may also lower the thresholds for attacks against democratic practices such as through disinformation”.⁷¹

‘Transparency’ is referenced three times in the document: “there may be unintended or unforeseen consequences of using AI as a result of biased or manipulated data, lack of transparency, misuse or hostile use”,⁷² “a cross-cutting theme should be sustainable AI, meaning that AI applications should be ethical, safe, secure, reliable and transparent”⁷³ and on page 8, when referencing algorithms: “The use of AI algorithms must be transparent and comprehensible”.

In Sweden, openness and transparency are touted as vital public values,⁷⁴ and it is discouraging that there is no mention of openness, whatsoever. Transparency is the most frequented of the three cultural values, however it is not developed in-depth. The strategic policy is lacking in inclusion of the three cultural values, however, its most striking statement might be in relation to use of and managing data necessary for AI: “such frameworks must balance fundamental needs for privacy, ethics, trust and social protection with access to the data needed to realise the potential of AI”.⁷⁵

‘Trust’ is mentioned twice in the 12-page document:

There may be unintended or unforeseen consequences of using AI as a result of biased or manipulated data, lack of transparency, misuse

70 Sweden’s Ministry of Enterprise and Innovation (2018), p. 5.

71 Ibid., p. 8.

72 Ibid., p. 8.

73 Ibid., p. 4.

74 Swedish Institute (2016).

75 Ministry of Enterprise and Innovation (2018), p. 10.

or hostile use. This may lead to discrimination, loss of trust, financial damage and consequences for the functioning of democracy.⁷⁶

And

Appropriate frameworks of principles, norms, standards and rules are therefore important prerequisites if Sweden is to realise the benefits of AI in society. Such frameworks must balance fundamental needs for privacy, ethics, trust and social protection with access to the data needed to realise the potential of AI.⁷⁷

Notably, Sweden has experienced government debacles similar to Denmark's Gladsaxes episode. For example, recent litigation cases have been launched against the Swedish municipality of Trelleborg, due to its potential overreach in attempts at automating civil services. Already in 2017, the municipality had prided itself for piloting programmatic decision-making algorithms as a tool for its civil servants when handling social benefit applications.⁷⁸ Moreover, the country has seen nationwide debates on the data management of its public authorities in recent years, with a news story about potential data leaks due to mismanagement in its national authority for transport (Transportstyrelsen) in 2017, news about data leaks from healthcare provider 1177 breaking in 2019, and a national debate on public-service data hosting in the wake of the U.S. American Cloud Services Act in 2019.

To conclude the analysis of the Swedish policy document, it therefore seems appropriate to state how more competently written the Danish, Norwegian and Finnish policy documents are. In stark contrast to the Danish, Finnish and Norwegian report, the Swedish report lacks clear direction in how Swedish government organisations will ensure openness, transparency and trust in adopting AI throughout society. The document

⁷⁶ Ibid., p. 4.

⁷⁷ Ibid., p. 10.

⁷⁸ Lind & Wallentin (2020).

is quite concise and does not set forward a strong set of policies and principles for which Swedish society can compete on a global scale.

3. Analysis of the shared initiatives

THE ANALYSIS OF the four Nordic national AI documents found three cultural values (openness, transparency and trust) are mentioned and upheld to various degrees, whether being explicitly mentioned, or present through related themes (i.e., ethics, autonomy, privacy and democracy) being examined. In summary, explicit or related themes of openness, transparency and trust have clear impact on the national strategic AI policies for Denmark, Finland, Norway and Sweden.

Not surprising, the Nordic countries have identified sharing of data as a competitive advantage, cooperating through the Nordic Council of Ministers:

A working group has been formed to identify datasets that can be exchanged between Nordic countries and create added value for Nordic enterprises – public and private alike – while still respecting the ethical aspects and the trust and values particular to the Nordic countries.⁷⁹

Building on the Nordic platform for data sharing (SECTION 3.3.1), Norway's policy document states:

Data can be regarded as a renewable resource. Sharing data with others does not mean that one is left with less data. In fact, the value of data can increase when shared because it can be combined with other types of data that can offer new insights or be used by organisations with the expertise to use the data in new and innovative ways.⁸⁰

⁷⁹ Nordic Council of Ministers (2020), p. 14.

⁸⁰ Norway's Ministry of Local Government and Modernisation (2020), p. 13.

Might the Nordics utilise their data sharing platform as a new sustainable business model? The noted data sharing is sustainable in that (1) data can be “re-used”, and (2) the energy savings seen as result of not needing to generate new data. However, not all sustainable initiatives like data sharing are without concerns, according to Norway’s policy document: “The costs of making datasets genuinely reusable must be weighed against the benefit to research communities and society”.⁸¹ However, data sharing or pooling has benefits for the Nordics, allowing a competitive advantage versus larger national or even private companies with massive amounts of data. One example of data sharing in Norway includes daily production figures from all oil wells in the Norwegian sector, which is then published on the Norwegian Petroleum Directorate website.⁸²

In addition to data sharing, the use of synthetic data for ensuring privacy is important, as well. Synthetic data is a privacy-enhancing process for making realistic synthetic data with properties of the original data set, without many of the privacy issues related to real data.⁸³ Norway’s policy document states:

Synthetic data can in many cases be an alternative to identifiable data or anonymised data. If synthetic datasets can be produced with the same features as the original dataset, they can be used to train algorithms or be used as test data. This means that even datasets which normally would be considered sensitive could be made openly accessible for use in research and innovation.⁸⁴

The use of data sharing and synthetic data opens many possibilities for enhancing privacy and sustainable data usage.

81 Ibid., p. 15.

82 Hass et al. (2017).

83 El Emam, Mosquera & Hoptroff (2020).

84 Norway’s Ministry of Local Government and Modernisation (2020), p. 17.

4. Discussion and policy recommendations

4.1. Public AI education initiatives

While the Nordics are indeed capitalising on their shared values as a business advantage (i.e., sharing datasets cross-nationally within the Nordics), there is also an outward looking perspective of “*How might we help others benefit from our shared experiences?*”. One example is a global online course created by the Finnish higher education,⁸⁵ and Norwegian industry educating citizens about the basics of AI.⁸⁶ Requiring no previous knowledge of AI, the training course is an opportunity for citizens around the world to learn what AI can and cannot do, aiming to educate as many as possible about AI.⁸⁷ Incorporating this curriculum in primary education is highlighted in several of the documents, too. The sheer magnitude of the Nordics’ soft power can be viewed here, as this global AI education campaign might reinforce notions of Nordic openness, trust and transparency, in turn influencing positive attitudes towards conducting business in the Nordics. I believe adding “AI literacy” to media literacy and digital literacy initiatives is essential for ensuring technology comprehension for the mass society.

4.2. Each country is unique in how it applies shared cultural values

The influence of each Nordic country is clear in all respective policy documents; however, we should note each country does not just simply “copy and paste” in their policy document – each of the our Nordic nations analysed here has unique advantages, creating clear policy differences. For example, Norway humbly acknowledges difficulty in competing globally:

Although large countries such as the United States and China have resources with which neither Norway nor Europe can compete, there are areas where Norway and Europe have competitive advantages,

⁸⁵ Pekkarinen (2018).

⁸⁶ Norwegian Ministry of Local Government and Modernisation (2020).

⁸⁷ Ibid.

such as certain industrial applications of AI and trustworthy AI that takes data protection and ethical considerations into account.⁸⁸

Later, the document then recognises how these constraints can become a competitive advantage:

As a small country, Norway does not have the capability to build knowledge and expertise to high international standards across the full spectrum of AI. Nevertheless, the quality and scope of our national expertise must be sufficient to exploit the technologies and innovations that emerge internationally. Another goal must be to leverage our position as a nation with a digitally advanced population and business sector in order to take the lead in applying AI, not least in industry.⁸⁹

Further, “Norway can take a leading position in applying artificial intelligence, particularly in areas where we already are well positioned and have strong business and research communities, such as health, oil and gas, energy, the maritime and marine industries and the public sector.”⁹⁰

Similarly, Denmark alludes to the same:

There is a global race within research in artificial intelligence, in which countries such as the US and China are investing massively in research in the technology. Even though Denmark does not measure up to large nations in terms of scope of research, Denmark stands strong as a research nation with good research environments within artificial intelligence. And Denmark is proportionally among the largest global investors in public research.⁹¹ However, its position can be advantageous and indeed influential: “Enhanced research efforts will also help ensure that Denmark can

88 Norway's Ministry of Local Government and Modernisation (2020), p. 36.

89 Ibid., pp. 36-37.

90 Ibid., p. 47.

91 Denmark's Ministry of Finance and Ministry of Industry, Business and Financial Affairs (2019), p. 44.

influence the development of artificial intelligence according to Danish values”.⁹²

Comparable notions are found in the Swedish policy document,⁹³ and the Finnish document:

It is not realistic for us to compete... with actors like China or the US. Still, Finland's limited resources and investments can be noteworthy, when targeted to strategically selected areas.⁹⁴

In short, *recognising your weaknesses can allow you to emphasise your unique advantages.*

4.3. National public policy should not lack clarity and depth

While some national policies are lacking in clarity and specificity (i.e. Sweden), there are documents lucidly illustrating how cultural values can result in value-laden technology policy. By setting clear expectations and recognising the impact that cultural values can have in implementing technology policy, Denmark, Finland and Norway should be applauded for their efforts. Rather than minimising them, such as Sweden, they clearly state the value and necessity of applying openness, transparency and trust in AI policy. A disservice is provided to citizens, industry and researchers when policy documents are haphazardly published. The purpose of these documents is to inform these stakeholders how a nation might address challenges of AI, what issues are faced in implementing AI across different sectors, and to set forth guidelines for how the public sector will embrace AI. The value of government policy documents is diluted when they lack clarity, specificity and depth – especially when addressing flourishing foci, such as AI, where stakeholders need clarity as to how the technology works, what changes might occur in society due to AI, and how cultural values

92 Ibid.

93 Sweden's Ministry of Enterprise and Innovation (2018), pp. 8-9.

94 Ministry of Economic Affairs and Employment of Finland (2017), p. 80.

will be sustained in light of the technology. Clearly, Sweden missed an opportunity, while the governments of Denmark, Finland and Norway delivered what industry, citizens and researchers needed – reassurance that AI will not erode openness, transparency and trust, but rather AI will provide new, challenging opportunities to reinforce and build upon these societal values.

Given that Sweden’s economy is the largest among the Nordic nations,⁹⁵ the overall vagueness and brevity of the Swedish policy document is alarming. The guide’s self-described goal is to “be the world’s leader in harnessing the opportunity offered by digital transformation”.⁹⁶ It also claims that, “by international standards, Sweden is in the vanguard”,⁹⁷ and argues that “Sweden can take the lead in ethical, safe, secure and sustainable use of AI by actively working on this issue nationally and promoting it internationally”.⁹⁸ Unfortunately, the guide accomplishes little more than the making of vague references to the power of AI and grandiose claims of how it will change Swedish society. In stark contrast to the Danish, Finnish and Norwegian policies, the Swedish policy lacks clear direction in how Swedish government organisations will ensure openness, transparency and trust in adopting AI throughout their society. Also absent are any strong policies and principles by which Swedish society can compete on a global scale. Nonetheless, it is useful to see how the values of openness, transparency and trust appear (or do not appear) in this document.

95 Nordic Statistics Database (2020).

96 Sweden’s Ministry of Enterprise and Innovation (2018), p. 4.

97 Ibid.

98 Ibid., p. 8.

5. Conclusion

WHILE AI HAS already begun to be a hands-on challenge for many Nordic administrations, as is shown in the Danish Gladsaxe example and the Swedish Trelleborg example (above), the general takeaway of the national policy documents is that, in these, AI is largely discussed in both a very general and in a future tense. The policy documents can, as such, be read as representations of the conceptual world (general ways of thinking about AI in relation to society) that the documents are samples of. In other words, the policy documents might give insight into the deeper structure of how thoughts about AI in relation to society are constructed.

My policy document review has found that while openness (a core societal value in the Nordics) is largely absent in the policy documents, the values of transparency and trust manifest explicitly or through related concepts. Regarding ambitious concepts such as interpersonal and institutional trust, recent examples have appeared in Nordic countries, as in the municipalities of Gladsaxe (Denmark) and Trelleborg (Sweden), that might actually serve to attenuate public trust in AI, as these municipalities were found to have employed AI-based systems for providing state services, where important decisions were made from potentially spurious or even privacy-encroaching data. Even more troubling would be those cases where data is potentially leaked, due to neglect or overly convoluted management systems, such as in the nationally well-known Transportstyrelsen and 1177 scandals in Sweden.

Each Nordic country presents vastly different policy documents, with one being a clear example of “what not to do” (i.e. Sweden). The government policies of Denmark, Finland and Norway should be generally commended for tackling complex technical issues and informing how AI will potentially impact society – including how citizens will interact with government, and receive services from those very governments.

References

- Abbass, H.A., Petraki, E., Merrick, K., et al. (2016). Trusted Autonomy and Cognitive Cyber Symbiosis: Open Challenges. *Cognitive Computation*; 8: 385-408. DOI: 10.1007/s12559-015-9365-5.
- Alfter, B. (2019). *Automating Society – Denmark*. Algorithm Watch.
- Bainbridge, J., Goc, N. and Tynan, L. (2008). *Media and journalism: New approaches to theory and practice*. Oxford, UK: Oxford University Press.
- Berggren, N. and Bjørnskov, C. (2011). Is the importance of religion in daily life related to social trust? Cross-country and cross-state comparisons. *Journal of Economic Behavior & Organization*; 80: 459-480. DOI: <https://doi.org/10.1016/j.jebo.2011.05.002>.
- Bunyan, T. (1999). *Secrecy and Openness in the EU*. Kogan Page.
- Byrne, C.S. and Sommer, J. (2019). Is the Scandinavian Digitalisation Breeding Ground For Social Welfare Surveillance? *DataEthics*. Denmark.
- Curtin, D. and Meijer, A.J. (2006). Does transparency strengthen legitimacy? *Information Policy*; 11: 109-122. DOI: 10.3233/IP-2006-0091.
- Denmark's Ministry of Finance and Ministry of Industry, Business and Financial Affairs. (2019). National Strategy for Artificial Intelligence. The Danish Government. https://eng.em.dk/media/13081/305755-gb-version_4k.pdf, accessed October 20, 2020.
- El Emam, K., Mosquera, L. and Hoptroff, R. (2020). *Practical Synthetic Data Generation*. Sebastopol, CA: O'Reilly Media.
- Eurobarometer. (2018). *Special Eurobarometer 471: Summary: Fairness, inequality and intergenerational mobility*. European Commission.
- Globaliseringsrådet. (2008). *Sweden in the world: Report*. Stockholm: Utbildningsdepartementet.
- Götz, N. (2015). The concept of openness: Promise and paradox. In: Götz N and Marklund C (eds.) *The paradox of openness: Transparency and participation in Nordic cultures of consensus*. Leiden, Netherlands: Brill, pp.10-26.
- Grimmelikhuijsen, S., Porumbescu, G., Hong, B., et al. (2013). The Effect of Transparency on Trust in Government: A Cross-National Comparative Experiment. *Public Administration Review*; 73: 575-586. Article. DOI: 10.1111/puar.12047.
- Hass, J.L., Haugen, J., Nielsen, V.O., et al. (2017). *Oil and gas statistics: The Norwegian experience*. Oslo, Norway: Statistics Norway.
- Hawkins, J. (2017). Textual analysis. In: Allen M, (ed.) *The SAGE Encyclopedia of Communication Research Methods*. Thousand Oaks, California: SAGE Publications, p. 1754-1756.
- Hofstede, G. (2019). Compare Countries - Hofstede Insights, <https://www.hofstede-insights.com/product/compare-countries/>, accessed February 28, 2020.
- Hood, C. (2006). Transparency in Historical Perspective. In: Hood C and Heald D (eds) *Transparency: The Key to Better Governance?* Oxford: Oxford University Press, pp.3-23.
- Jorgensen, O. (2014). *Access to Information in the Nordic Countries: A comparison of the laws of Sweden, Finland, Denmark, Norway and Iceland and international rules*. Göteborg: Nordicom, p.40.
- Lind, K. and Wallentin, L. (2020). Central authorities slow to react as Sweden's cities embrace automation of welfare management, <https://algorithmwatch.org/en/story/trelleborg-sweden-algorithm/>, accessed March 20, 2020.

Marozzi, M. (2015). Measuring Trust in European Public Institutions. *Social Indicators Research*; 123: 879-895. Article. DOI: 10.1007/s11205-014-0765-9.

Nordic Council of Ministers. (2020). About the Nordic Council of Ministers, <https://www.norden.org/en/information/about-nordic-council-ministers>, accessed March 4, 2020.

Ministry of Economic Affairs and Employment of Finland. (2017). Finland's Age of Artificial Intelligence: Turning Finland into a leading country in the application of artificial intelligence Objective and recommendations for measures. In: Employment MoEa, (ed.). Helsinki. https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf, accessed October 20, 2020.

Ministry of Local Government and Modernisation (2020) National Strategy for Artificial Intelligence. Norway. [Kommunal-og moderniseringsdepartementet (2020) Nasjonal strategi for kunstig intelligens].

Nordic Statistics Database. (2020). NAACo1: National accounts, supply and demand, million euro by reporting country, national accounts indicator, unit and time, https://pxweb.nordicstatistics.org/pxweb/en/Nordic%20Statistics/Nordic%20Statistics__Economy__National%20accounts/NAACo1.px/table/tableViewLayout2?rxid=6do2d604-759c-499a-997c-4e353f47962a, accessed September 2, 2020.

Nordic Council of Ministers. (2020). *Nordic cooperation on data to boost the development of solutions with artificial intelligence*. <https://pub.norden.org/nord2020-042/#>, accessed October 20, 2020.

Norway's Ministry of Local Government and Modernisation. (2020). National Strategy for Artificial Intelligence. Norway. [Kommunal-og moderniseringsdepartementet (2020) Nasjonal strategi for kunstig intelligens].

Norwegian Ministry of Local Government and Modernisation. (2020). National Strategy for Artificial Intelligence.

OECD. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*.

OECD. (2019). Openness and Transparency - Pillars for Democracy, Trust and Progress, <https://www.oecd.org/fr/etatunis/opennessandtransparency-pillarsfordemocracytrustandprogress.htm>, accessed March 5, 2020.

Organisation for Economic Co-operation and Development. (2019). Trust in government - Openness, <http://www.oecd.org/gov/trust-openness.htm>, accessed March 4, 2020.

Ortiz-Ospina, E. and Roser, M. (2020). Trust, <https://ourworldindata.org/trust>, accessed March 10, 2020.

Pekkarinen, A. (2018). Artificial intelligence needs people: Three reasons to learn the basics now. <https://www.helsinki.fi/en/news/data-science-news/artificial-intelligence-needs-people-three-reasons-to-learn-the-basics-now>, accessed March 10, 2020.

Robinson, S.C. (2018). Factors predicting attitude toward disclosing personal data online *Journal of Organizational Computing and Electronic Commerce*; 28: 214-233. DOI: 10.1080/10919392.2018.1482601.

Robinson, S.C. (in press). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for Artificial Intelligence (AI). *Technology in Society*, 63, 00-00. <https://doi.org/10.1016/j.techsoc.2020.101421>

Schlagwein, D., Conboy, K., Feller, J., et al. (2017). "Openness" with and without Information Technology: a framework and a brief history. *Journal of Information Technology*; 32: 297-305. DOI: 10.1057/s41265-017-0049-3.

Smith, P.B., Andersen, J.A., Ekelund, B., et al. (2003). In search of Nordic management styles. *Scandinavian Journal of Management*; 19: 491-507. DOI: 10.1016/S0956-5221(03)00036-8.

Sorgenfri Kjaer, J. (2019). Gladsaxe indstiller arbejdet med omstridt overvågning af børnefamilier. *Politiken*, February 6. Available at: <https://politiken.dk/indland/art7023935/Gladsaxe-indstiller-arbejdet-med-omstridt-overvågning-af-børnefamilier>, accessed October 26, 2020

Sønderskov, K.M. and Dinesen, P.T. (2016). Trusting the State, Trusting Each Other? The Effect of Institutional Trust on Social Trust. *Political Behavior*; 38: 179-202. DOI: 10.1007/s11109-015-9322-8.

Stevenson, B. and Wolfers, J. (2011). Trust in Public Institutions over the Business Cycle. *American Economic Review*; 101: 281-287. Article. DOI: 10.1257/aer.101.3.281.

Stoyan, A.T., Niedzwiecki, S., Morgan, J., et al. (2016). Trust in government institutions: The effects of performance and participation in the Dominican Republic and Haiti. *International Political Science Review*; 37: 18-35. Article. DOI: 10.1177/0192512114534703.

Sweden's Ministry of Enterprise and Innovation. (2018). National Approach to Artificial Intelligence. The Government Offices of Sweden. <https://www.government.se/4a7451/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf>, accessed October 20, 2020.

Swedish Institute. (2016). Fact sheet: Openness and transparency, <https://sharingsweden.se/materials/openness/>, accessed March 4, 2020.

Torney-Purta, J., Henry Barber, C. and Richardson, W.K. (2004). Trust in Government-related Institutions and Political Engagement among Adolescents in Six Countries. *Acta Politica*; 39: 380-406. DOI: 10.1057/palgrave.ap.5500080.

Transparency International. (2016). Corruption Perceptions Index 2015, <https://www.transparency.org/cpi2015#downloads>, accessed March 5, 2020.

Vesa, J. (2015). Nordic Openness in Practice. *Nordicom Review*; 36: 129-142. DOI: 10.1515/nor-2015-0021.

Welch, E.W., Hinnant, C.C. and Moon, M.J. (2005). Linking Citizen Satisfaction with E-Government and Trust in Government. *Journal of Public Administration Research and Theory*; 15: 371-391. DOI: 10.1093/jopart/muio21.

AI policy in the Netherlands

More focus on practice than
principles when it comes
to trustworthiness

Katja de Vries

Katja de Vries is an Assistant Professor in Public Law at Uppsala University (Sweden). Her research is located at the intersection of law, philosophy and AI. Her current research focuses on the challenges that AI-generated content ('deepfakes' or 'synthetic data') poses to data protection, intellectual property and other fields of law. She is also affiliated to the Swedish Law and Informatics Research Institute (Stockholm), the Centre for Law, Science, Technology and Society (Brussels), and the Department of Sociology of Law (Lund).

Summary

THE NETHERLANDS IS eager to jump on the European AI wagon and become a frontrunner in exporting the European brand of human-centered AI. In terms of AI readiness, the Netherlands is a frontrunner in many respects, although the public funding for Research and Development (R&D) in AI has been criticised for having a rather shaky basis. In terms of AI that aligns with fundamental rights and values, the Netherlands has strong preconditions. Yet, cases like *CAF-11* and *SyRI* have also shown that these beneficial preconditions do not guarantee that the right balance is struck in practice. In its national AI strategy, the Netherlands does not follow the policy documents of the High-Level Expert Group on Artificial Intelligence (AI HLEG), with little mention of trustworthiness. The three central documents

constituting the national AI strategy (the *Strategic Action Plan for Artificial Intelligence*, SAPAI, published in October 2019, and two governmental letters) are sceptical of abstract concepts and vague ethical guidelines, instead focusing on how guidelines can be operationalised pragmatically within the national Dutch context and given bite through practical tools, legislation and supervision. However, while pragmatism is sympathetic, the Dutch AI strategy might have benefited from some more overarching vision. Dealing with the national strategy in three separate documents brings along a strategy that is somewhat dispersed, and also gives the impression that the fundamental values and rights are purposefully kept outside the upbeat glossy SAPAI report. Given the recent *CAF-11* and *SyRI* cases in the Netherlands, it is clear that the public sector has had a difficult time in finding a middle ground between governmental efficiency and protection of fundamental rights.

1. Introduction and chapter structure

BEFORE DIVING INTO the question of how the Dutch AI strategy conceptualises *trustworthy AI* (in section 4 and 5), I will give a general introduction about (2) the level of AI readiness in the Netherlands, and present (3) some recent cases from the Netherlands illustrating the struggle of the public sector to balance governmental efficiency and respect for fundamental rights in automated decision making. Given that the concept of trustworthy AI is not a disembodied notion, it is important to place it in this broader setting. As stated in the conclusions (6), the creation of trustworthy AI does not merely require a context-aware, practical operationalisation of the concept, but also requires financial, institutional and cultural support.

2. The Netherlands – a frontrunner in AI readiness?

AS HAS BEEN widely noted, Europe is struggling to catch up in terms of AI with the US and China.¹ Within Europe, however, The Netherlands

¹ Khari Johnson (2019).

has a reputation of belonging to the forerunners in terms of AI readiness. Its “entrepreneurial culture, flat organizations and growth of innovative companies”,² for which it was recently praised by the World Economic Forum (WEF), are beneficial for any economic activities, including those in the field of AI. One of the conditions that are particularly beneficial for AI readiness is the data connectivity, for which the Netherlands has been ranked as one of the leading countries in the world.³ In the Netherlands 2.6 percent of GDP is produced by the digital- and AI-based ICT sector, which is significantly higher than the European average of 1.66. This results in the Netherlands ranking among the high achieving countries like the US (3.3 percent), Finland (3.0 percent), Sweden (2.8 percent) and China (2.2 percent). It should, however, be kept in mind that in absolute numbers the digital- and AI-based ICT sector of large countries like the US and China is much larger. In a recent ranking, looking at 172 countries in terms of *Government AI Readiness*,⁴ the Netherlands ranked 9th place, behind the US (1st place) but well before China (19th place). Also in terms of AI research the Netherlands is doing well, ranking 13th place in 2019 in the top 20 of leading countries.⁵

Some aspects of AI readiness could be improved in the Netherlands. For example, although it scores above the European average, in human skills capital (the available workforce possessing AI skills), the Netherlands is behind European frontrunners such as Finland, Sweden and Germany.⁶ Another improvement could be with regards to state investments and if they are sufficiently large, for example: France has pledged to spend

2 NL Times (2019); See also the ranking of the Netherlands in the *European innovation scoreboard 2020* as belonging to the top 5 of most innovative countries within the EU, “where performance is above 125% of the EU average. The Innovation Leaders are Denmark, Finland, Luxembourg, the Netherlands, and Sweden”, European Commission (2020a), p. 13.

3 Tralac (2016).

4 Oxford Insights and the International Research Development Centre (2020).

5 Chuvpilo (2019). It is noteworthy that the differences in publication indices are quite large though is this ranking: 1. United States — 1260.2; 2. China — 184.5; 3. United Kingdom — 126.1; 4. France — 94.3; 5. Canada — 80.3; 6. Germany — 64.57; Switzerland — 59.3; 8. Japan — 49.4; 9. South Korea — 46.8; 10. Israel — 43.3; 11. Australia — 27.0; 12. India — 17.1; 13. Netherlands — 15.3; 14. Singapore — 13.2; 15. Denmark — 12.2; 16. Italy — 11.5; 17. Sweden — 11.3; 18. Russia — 10.6; 19. Finland — 9.6; 20. Austria — 7.4.

6 Oxford Insights and the International Research Development Centre (2020), p.40.

1.5 billion EUR of public funding on AI by 2022,⁷ and Germany injected 500 million EUR in 2019 into AI and intends to have spent around 3 billion EUR on the implementation of their AI strategy by 2025. In July 2019, in preparation for the October 2019 release of the Dutch Strategic Action Plan for AI (SAPAI, see below), the Confederation of Netherlands Industry and Employers (known as VNO-NCW) released a position paper⁸ where they suggest that the Netherlands would need to invest 2 billion EUR⁹ in R&D on AI within private-public collaborations in the following 7 years. The rationale behind it is that approximately half of that amount would be raised by the private sector. The suggestion gained support in the government¹⁰ but not sufficiently¹¹ to make any hard commitments in the national strategy: there the commitments are limited to much smaller amounts. During the period 2019-2025 the Dutch Government intends to spend an annual base amount of 45 million EUR on AI.¹² In 2019 this base amount was increased by another 35 million EUR. In 2020 the Dutch government reserved another additional 23.5 million EUR for the period 2020-24. This results in a hypothetical government investment of around 64 million per year, which translates to a yearly per capita investment in AI R&D of 3,70 EUR. This is close to the European average (3,93 EUR), meaning higher than in the US (2,85 EUR) but below countries like France (4,45 EUR), Germany (5,16 EUR) and Singapore (16,72 EUR). However, if extra investments were made, The Netherlands could end up in the league of the relatively big spenders on AI.

What the final picture will be is not completely clear. The public funding for realising SAPAI is, to a substantial extent, a reshuffling of money (from existing innovation funding, etc.) and not allocation of new public funds. This results in uncertainty concerning the extent to which the

7 Rosemain & Rose (2018).

8 Taskforce AI (2019).

9 NOS (2019).

10 Nederlandse AI Coalitie (2020).

11 van der Starre & den Hollander (2019).

12 Ministerie van Economische Zaken en Klimaat. (2019); A slightly shortened English version is also available, Ministry of Economic Affairs and Climate Policy (2019a).

planned State funding will be realised. Particularly after the large economic strains following from the COVID-19 pandemic, there are worries that both public and private investments in AI in the Netherlands might become conservative.¹³ For example, in 2019, five Dutch multinationals (Ahold Delhaize, ING, KLM, NS and Philips) joined hands in the *KickstartAI* platform¹⁴ and committed to taking the lead in investing in AI: “boosting the AI ecosystem in the Netherlands by accelerating and promoting the development of AI technology and nurturing AI talent in the country”.¹⁵ However, COVID-19 led to unprecedented losses for transportation and airplane companies like NS (Dutch Railways) and KLM, putting into question the extent to which these commitments can and will be realised.

3. A history of questionable data use by the Dutch public sector

THE NETHERLANDS is generally known as a country governed by the rule of law and respect for fundamental rights.¹⁶ There is a general sentiment that these public values should also be upheld when algorithmic automated decision making (ADM), data analytics or AI-applications are involved, and that the Netherlands wants to be a forerunner in this respect too. One of many examples that can be named here is the Algorithm Register¹⁷ (currently still in *beta* phase) launched in September 2020 by the city of Amsterdam, that aims to give an “overview of the artificial intelligence systems and algorithms used by the City of Amsterdam.” While this pilot project is in itself laudable,¹⁸ and should not be judged harshly because of the early stage of its development, a closer look at some of the described algorithms provides insight into the difficult balance the public sector has to strike between

13 Deloitte (2020) & Deloitte AI Institute & Deloitte Center for Technology (2020).

14 Kickstart AI (2020).

15 Philips (2020).

16 See, for example, World Justice Project (2020): The Netherlands ranked 7th out of 128 countries with regards to respect for fundamental rights. The Netherlands is only preceded by four Scandinavian countries, Germany and Austria.

17 Amsterdam Algoritmeregister Beta (2020).

18 Johnson (2020).

effective governance (which includes making fraud detection more effective) and protection of fundamental rights of individual citizens. One of the algorithms that is described is “an algorithm that supports the employees of the department of Surveillance & Enforcement in their investigation of the reports made concerning possible illegal holiday rentals”. The information provided about the algorithm states that it “helps prioritize the reports so that the limited enforcement capacity can be used efficiently and effectively. By analysing the data of related housing fraud cases of the past 5 years, it calculates the probability of an illegal holiday rental situation on the reported address”.¹⁹ The information is extremely concise, making it impossible for anyone to be able to rig the system based on it. Is the information empowering for citizens? While it is unlikely that the information does much in terms of providing actionable transparency about the *workings* of the algorithm, it does provide transparency about the *existence* of the algorithm, which is good from a rule of law perspective (foreseeability) as well as from a governance perspective: it could have a deterrent effect for anyone considering renting out illegally. Two recent cases with regard to automated decision making (ADM) in the Dutch public sector have shown that the outcome of this balancing act between fundamental rights protection and governmental efficiency can easily lean too much into the direction of the latter, and that this might result in unjust and societally negative outcomes.

The first scandal, known as the CAF-11²⁰ case, concerned the Dutch tax office, where the CAF-antifraud team, colloquially known as the “cow-boy team”,²¹ used an algorithm to detect fraud regarding child benefits which, in combination with harsh legislation,²² led to financial tragedy and injustice for many of the affected families.²³ The scandal started in 2014, with origins going back as far as 2005, and is currently (autumn 2020) still an ongoing debate regarding the compensation to

19 City of Amsterdam Algorithm Register Beta (2020).

20 CAF is an acronym for *Combiteam Aanpak Facilitators* (Combiteam Handling Facilitators), an anti-fraud team working within the Dutch Tax office. See also Belastingdienst (2020).

21 Klein (2019).

22 Ministerie van Financiën (2020).

23 DutchNews (2020).

the potentially twenty thousand victims.²⁴ The algorithm in the CAF-11 case turned out to be both discriminatory (a double nationality allegedly being one of the factors contributing to being flagged as fraudulent)²⁵ and overly “efficient” (too many false positives). Its application created extensive human suffering: poor families that were misclassified as fraudulent ended up in a spiral of financial misery as they were forced to pay thousands, or even tens of thousands of euros, to the tax office. The CAF-11 case also had a political impact: in December 2019 Menno Snel, the State Secretary for Finance, resigned after receiving a parliamentary motion of no confidence for his handling of the CAF-11 case; and in May 2020, the Ministry of Finance pressed charges against five of its own high positioned civil servants - breaking with a hallowed tradition that merely politicians take the fall and bureaucrats are left untouched. In early 2020 a second scandal about the anti-fraud team was discovered: for almost 20 years the Tax Office had used a secret blacklist of citizens (in 2014 baptised the *Fraude Signalerings Voorziening* or FSV)²⁶ which were profiled as having a high likelihood of committing fraud. In 2019 this list contained around 180,000 people. After this second scandal the CAF-anti-fraud team was shut down.

The second case that throws an interesting light on the struggle of the Dutch public sector with ADM is the *SyRI case*. SyRI (System Risk Indication) was a legislation²⁷ that allowed the government to use a risk assessment tool “to detect various forms of fraud, including social benefits, allowances, and taxes fraud”.²⁸ Between 2008 and 2014 the risk tool emerged as a government practice that lacked a legal basis. In 2015

24 Rijksoverheid (2020).

25 Klein (2019).

26 Kleinnijenhuis (2020).

27 SUWI Act (2013). Section 65 SUWI Act (*Wet structuur uitvoeringsorganisaties werk en inkomen*) and Chapter 5a SUWI Decree (*Besluit van 20 december 2001 tot vaststelling van een algemene maatregel van bestuur ter uitvoering van de Wet structuur uitvoeringsorganisatie werk en inkomen, en in verband daarmee van enige andere socialezekerheidswetten*) together are known as the “SyRI legislation”. The SUWI Act was changed (2013) to give a legal basis to SyRI. This change was further elaborated in a change (2014) within the SUWI Decree.

28 SUWI Decree (2014).

a legal basis was created.²⁹ In an important ruling from February 2020,³⁰ The Hague District Court ruled that the SyRI legislation did not comply with Article 8 of the European Convention on Human Rights (ECHR), which protects the right to respect for private and family life, home and correspondence, because the SyRI legislation is “insufficiently transparent and verifiable”.³¹ This is particularly problematic for a tool like SyRI, which carries the risk that “discriminatory effects – unintentional or otherwise” (§ 6.91) will occur. The Court also concluded that SyRI “contains insufficient safeguards to protect the right to respect for private life.” (§ 6.95) The Court thus ruled that SyRI is unlawful, and has “no binding effect” (§ 6.112). In the meanwhile, the Dutch Government has proposed a new Act, *Wet gegevensverwerking door samenwerkingsverbanden* (WGS, Data Processing Partnership Act) which would provide a legal basis for public-private cooperation applying “to the processing of data, specifically when this processing is used for surveillance or investigation purposes, for example to prevent crimes, or to detect welfare fraud”.³² The law has an even wider scope than the SyRI legislation, with which it bears a significant resemblance, as pointed out in the critical report (published 29 April 2020) on this proposed Act from the Council of State.³³

Currently *CAF-11* and *SyRI* have become token cases in a movement of civil concern in the Netherlands for regulating the use of algorithms in decision making, particularly in the public sector.³⁴

29 See above: section 65 SUWI Act and chapter 5a SUWI Decree.

30 de Rechtspraak (2020b).

31 Ibid, para. 6.86.

32 van Til (2020).

33 Raad van State (2020).

34 See for example, NRC (2020).

4. The Dutch AI Strategy

4.1. How to read policy documents against the grain

On December 7th 2018, the European Commission presented a *Coordinated Plan on Artificial Intelligence* in which Member States were encouraged “to develop their national AI strategy by mid-2019, building on the work done at the European level”.³⁵ On October 8th 2019 the Dutch government sent three letters to the House of Representatives that were to serve as the basis for the national position with regard to AI:

- ▶ A letter from the Dutch State Secretary for Economic Affairs and Climate,³⁶ accompanied by a glossy 64-page report entitled *Strategic Action Plan for Artificial Intelligence* (abbreviated as SAPAI).³⁷
- ▶ A letter from the Dutch Minister for Legal Protection on safeguards against risks associated with data analytics by the government.³⁸
- ▶ A letter from the Dutch Minister of the Interior and Kingdom Relations on AI, public values and human rights.³⁹

These three letters are the core documents of the current Dutch strategy with regards to AI. The Dutch AI strategy is annually reviewed. The first review of the Dutch strategy was undertaken in the annual governmental progress report on digitisation, which includes a review of SAPAI (June 2020).⁴⁰ The annual review has not brought any changes to the strategy, and the three governmental letters from October 2019 continue the main point of reference: the Dutch government consistently refers to them as a point of departure for all other AI policy in the Netherlands.⁴¹

35 European Commission (2018), see also chapter 1 in this volume.

36 Ministry of Economic Affairs and Climate Policy (2019b).

37 Government of the Netherlands (2019).

38 Minister for Legal Protection (2019).

39 Minister of the Interior and Kingdom Relations (2019).

40 Ministerie van Economische Zaken en Klimaat (2020).

41 See, for example, the annex to Parliamentary Papers 2019-2020: Government assessment of White Paper on Artificial Intelligence (2020). In this document the Dutch government assesses the European Commission’s digital strategy for 2020-2025 (published on the 19 February 2020) from the point of view of the Dutch AI strategy. The document also offers a good overview in English of the Dutch strategy.

Staying true to the genre of policy writing, these three governmental letters are filled with good intentions, great promises and long sentences. In order to give a critical assessment of these documents, it is pivotal to not merely look at their actual content, but also ask questions with regards to its *creators* (who wrote these texts? why these authors?), *organisation* (why three letters and not just one big report?), *reception* (what reactions did it generate and from whom?), *context* (are these texts explicitly, or implicitly, reacting to events from the societal and political context?) and *effects* (what impact?).

In the remainder of this contribution I aim to give a concise overview of the Dutch strategy with regards to AI, focusing on how trustworthiness of AI is conceptualised. I will first present some of the salient characteristics of the three governmental letters and the SAPAI report, using the aforementioned critical questions as a guideline. I juxtapose and compare the Dutch approach with the European Commission's approach on "trustworthy AI".⁴² In the final part, I place the Dutch approach on trustworthy AI in the Dutch *context* discussed in the two opening sections of this chapter.

4.2. The three tracks of Dutch AI strategy – is there a dominant track in the multifaceted strategy to AI in the Netherlands?

AI affects almost every aspect of life, which unsurprisingly leads to uncertainty as to which state agency is responsible for it. This ambiguity is reflected in the fact that the three governmental letters are authored by different ministries (see above) and that the SAPAI report itself was a collaborative governmental effort. Included in the writing of the report, headed by the Ministry of Economic Affairs and Climate Policy (more specifically the *digital economy division of the directorate-general for enterprise and innovation*), was the Ministry of Justice and Security, the Ministry of Education, Culture and Science, the Ministry of Social Affairs and Employment, and the Ministry of the

42 See: High-Level Expert Group on AI (2019a); High Level Expert Group on AI (2019b) & European Commission (2020b).

Interior and Kingdom Relations participated. On the same day as the three governmental letters were released, the *Netherlands AI Coalition* (NL AIC) was launched. This is a public-private partnership

... in which the government, the business sector, educational and research institutions, as well as civil society organisations collaborate to accelerate and connect AI developments and initiatives. The ambition is to position the Netherlands at the forefront of knowledge and application of AI for prosperity and well-being. We are continually doing so with due observance of both the Dutch and European standards and values. The NL AIC functions as the catalyst for AI applications in our country. One of the main goals is to have impactful AI in at least ten economic and societal sectors within the coming three years. This is where the application of AI takes place and where we create our economic impact and tackle societal challenges.⁴³

TO THE RIGHT: FIGURE 1: From the SAPAI report.⁴⁴

The Dutch clearly struggled with the question of how to organise the multiplicity of stakes, interests, perspectives and approaches to AI in one coherent AI strategy. The Dutch AI strategy proposed in the SAPAI report is organised in three tracks, summarised in Figure 1: *Capitalising on societal and economic opportunities* (track 1), *Creating the right conditions* (track 2), and *Strengthening the foundations* (track 3).

In the introduction to the SAPAI the three tracks are presented:

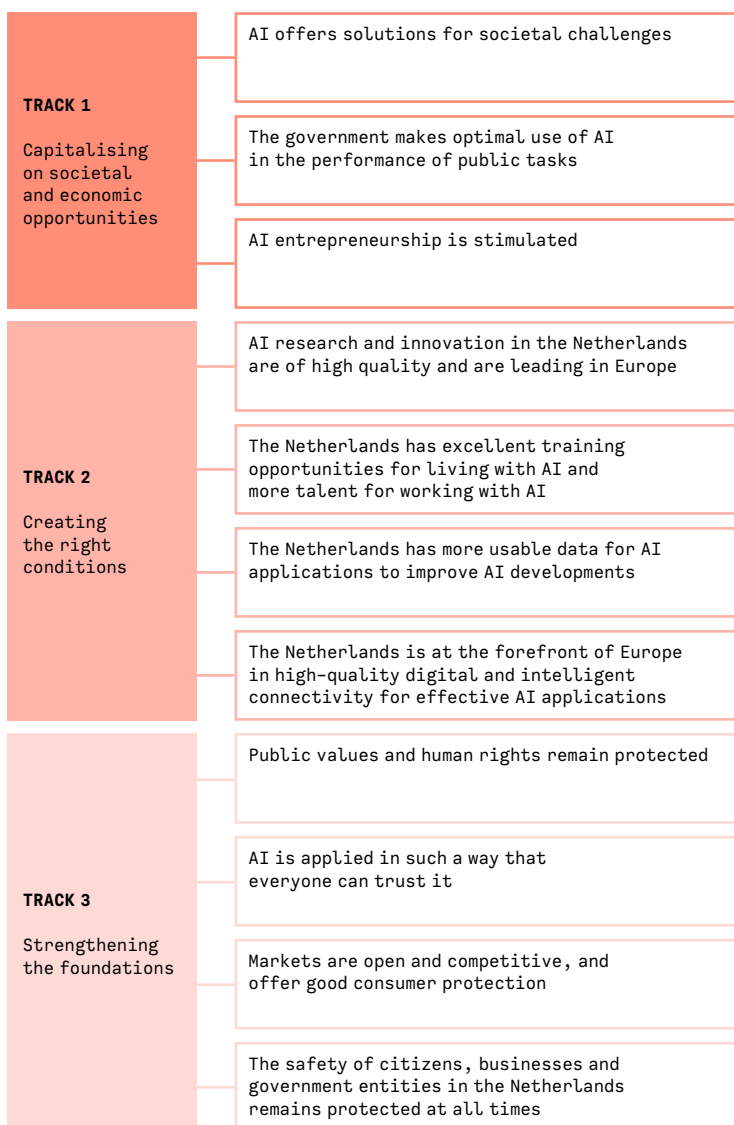
TRACK 1 is that we must capitalise on societal and economic opportunities. [...] [Companies] will determine whether the Netherlands leads the way in AI or ends up following other countries in this area, and whether AI applications will really benefit the Dutch economy and Dutch society. Companies' added value comes from their ability to respond both to societal challenges and market demand. As a

43 NL AI Coalition (2020).

44 Ministry of Economic Affairs and Climate Policy (2019a), p. 6.

STRATEGIC ACTION PLAN FOR ARTIFICIAL INTELLIGENCE

The Netherlands is able to capitalise on AI's societal and economic opportunities, as well as to safeguard the public interests of AI, thus contributing to prosperity and well-being



result, the government is developing this track in PPPs [public-private partnerships], particularly with the Dutch AI Coalition, while calling on companies and organisations to join in these efforts.

TRACK 2 intends to arrange the required prerequisites for a favourable AI climate in the economy and society at large. These prerequisites include the right knowledge, skills and training: top-quality scientific AI research as well as applied research that businesses and professionals can use. They also include usable data and high-quality and intelligent connectivity. In this context, we are investing in research programmes, increasing access to innovation funding for start-ups, investing in training, promoting data-sharing, and further developing digital connectivity with and for AI.

TRACK 3 is about ‘Strengthening the foundations’. This track concerns the protection of citizens’ fundamental rights as well as appropriate legal and ethical frameworks. As a result, people and companies will feel confident that AI will be used with care. It is important as well that markets remain open and competitive, and that national security is safeguarded in all AI developments. To this end, the necessary legislation and regulations are already in force or in the making. The government is also monitoring AI developments and making efforts to ensure the use of ethical guidelines (European and otherwise) for AI applications by companies and public organisations.⁴⁵

The Dutch report takes pride in its balanced, three track approach, in which economic and societal initiatives, research, technological infrastructure, ethical and legal foundations intertwine. However, although the authors of both the governmental letters and the SAPAI report would probably deny this, there are several reasons to suspect that the numbering of the tracks also reveals something of an ordering in terms of importance: that the entrepreneurial (notably in the form of private-public collaborations, PPCs) track, dominates over the others. A

45 Ministry of Economic Affairs and Climate Policy (2019a), p.7.

first indication can be found in the fact that the Ministry of Economic Affairs is the *author* of the letter presenting the SAPAI report.

A second indication can be found in the *reception* of the SAPAI report: two reactions dominate. Firstly, from an entrepreneurial and commercial perspective, there are mainly supportive reactions for the governmental support for AI. Secondly, from a research perspective, there is a fear that the big promises from the report will not be represented in increased financial support for research in AI. This is both because much of the funding is likely to come from private actors in PPPs, and because State funding of AI research may not materialise or be used as expected. Shortly after the publication of the SAPAI report, the scientific director of the Dutch AI innovation centre ICAI⁴⁶ complained in one of the main Dutch newspapers that it is difficult to keep or attract top researchers when state investments into AI are significantly higher in neighbouring countries like France and Germany.⁴⁷

The third track of the SAPAI report (detailed in Figure 2) hardly generated any response from civil society or academia.⁴⁸ This brings me to the third indicator for the subordinated role of research and fundamental values in comparison to the capitalisation of AI opportunities: the textual *organisation* and presentation of the Dutch strategy. Why not one report, but three letters? One can argue that this simply is a governmental division of labour. However, the SAPAI report itself is already a collaborative writing effort. So why not integrate the letters on *safeguards against risks associated with data analytics by the government* and *AI, public values and human rights* into track 3 of the report? Is it merely because track 3 in the report would take up a disproportionately large part of the report? Or might it be because the glossy format and upbeat tone of the SAPAI report would be destroyed by too many critical reflections? The governmental letter on Legal Protection on safeguards against risks associated

46 Innovation Center for Artificial Intelligence (2020).

47 Bronzwaer (2019a).

48 See for an exception, this (mainly critical) analysis of SAPAI: van den Hoven van Genderen (2019).

with data analytics by the government gives a positive twist to it,⁴⁹ by stating that this separate treatment is done exactly because the role of fundamental values in AI deserves special and extended attention:

SAPAI deals in track 3 shortly with the effect of AI on public values. Because of the complexity and potential significance of the effects of AI on public values and human rights, the government has chosen to devote additional attention to policy in this respect in its governmental letter on AI, public values and human rights. The same goes for the current letter, that deals more specifically with potential safeguards against the risks following from the use of algorithms and data analytics by the government. (author's translation)

Track 3 in the SAPAI report includes a long list of ongoing and new actions to stimulate aspects of trustworthy AI, but does not provide any innovative conceptual considerations. In this regard the SAPAI report simply refers to the 'Ethics Guidelines for Trustworthy AI' of the High-Level Expert Group on Artificial Intelligence (AI HLEG). In an informational chart (see figure 2) the report lists actions that would increase trust in AI in the Netherlands, which amount to (1) Dutch companies and public organisations participating in the pilot phase of the AI HLEG'S ethics guidelines and (2) research into responsible development, use and supervision of algorithms.

With regard to the first action, the SAPAI report states that, with the AI HLEG's ethics guidelines in place, "it is now up to the market to decide how to implement and apply these guidelines".⁵⁰ Concrete actions to support this include: an investigation of the Ministry of Economic Affairs and Climate Policy into the use and risk management of algorithms in different sectors, contributions of the NEN Standards Committee to good practices and standards for AI, experimentation by the Ministry of the Interior and Kingdom Relations with instruments such as the AI Impact Assessment and quality marks in governmental use of algorithms, and

49 Minister for Legal Protection (2019).

50 Ministry of Economic Affairs and Climate Policy (2019a), p.43.

collaboration between supervisory authorities to see if there are persistent blind spots in the supervisory landscape with regard to AI.

With regard to the second action, the report says that the Dutch government is “encouraging”⁵¹ such research – a wording vague enough to avoid any real commitment. Later in the report this governmental “research encouragement” is further elaborated through two specific actions: (1) a call of the Dutch Organisation for Scientific Research worth 2.3 million EUR, on explainable, socially aware and responsible AI (closing date November 5th 2019), (2) research calls for public-private partnerships into the responsible use of AI and the transparency and explainability of algorithms.

ON THE NEXT PAGE: FIGURE 2. From the SAPAI report.⁵²

While the actions in themselves are laudable, they do not stand out as building on an abundance of vision or financial commitments.

In summary, the first track (*capitalising on social and economic opportunities*) of the three tracks of Dutch AI strategy seems to be the one dominating the other two despite the fact that the report does not explicitly talk of such hierarchy. This aligns with a reaction of the Director of the Foundation for Digital Infrastructure in the Netherlands to the SAPAI report,⁵³ who considered it to be representative of Dutch economic policy in general: “The market will fix it, and if it doesn’t turn out to be perfect, you will have to subsidize a bit”.⁵⁴

I now turn to the two governmental letters (*Legal Protection on safeguards against risks associated with data analytics by the government* and on *AI, public values and human rights*) that elaborate on the “strengthening the foundations”-track. Do these letters give further indications as to how the Dutch AI strategy elaborates the notion of trustworthy AI?

51 Ibid.

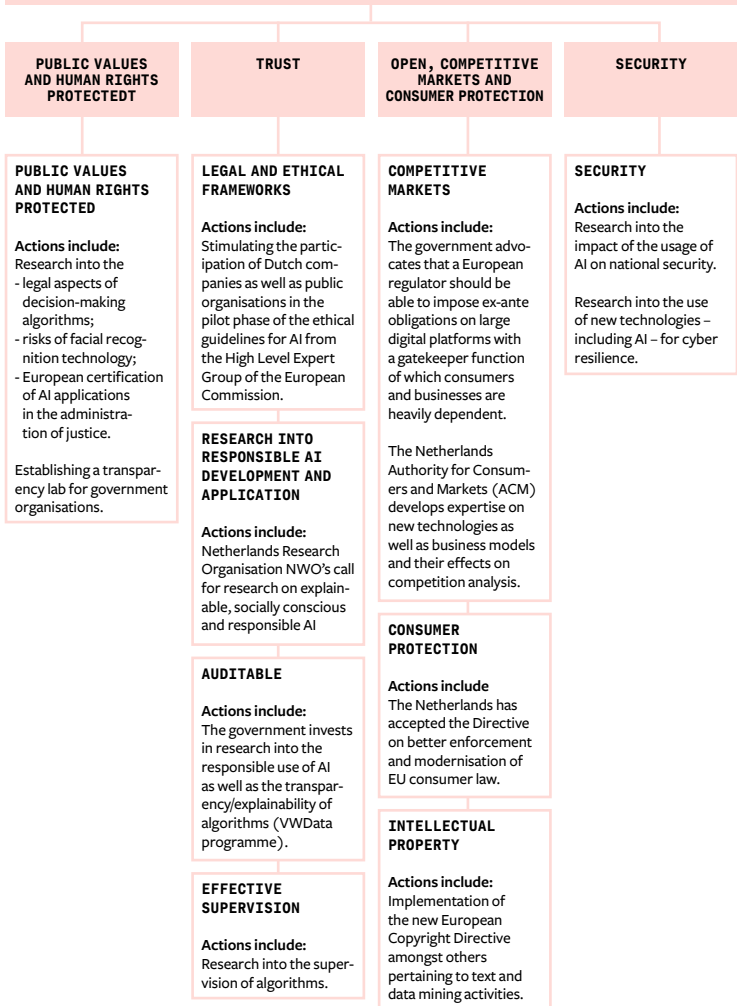
52 Ministry of Economic Affairs and Climate Policy (2019a), p. 40.

53 Stichting Digitale Infrastructuur Nederland (2020).

54 *Translation mine (KdV)*. Bronzwaer (2019b).

TRACK 3

Strengthening the foundations



From the SAPAI report.

5. Two governmental letters that give more substance to trustworthy AI?

WHILE THE DISCUSSION of track 3 in the SAPAI report is rather shallow, the discussion gets substantially more thorough in the governmental letters written by the Minister for Legal Protection on *safeguards against risks associated with data analytics by the government*,⁵⁵ and by the Minister of the Interior and Kingdom Relations on *AI, public values and human rights*.⁵⁶ Unfortunately these letters are only available in Dutch, which makes them inaccessible for large parts of the international community. It is therefore important to discuss their content in some more detail.

I begin by discussing the letter on safeguards in the use of data analytics by the government. This 13-page letter builds on other policy documents (mainly from 2018) and on a report published in March 2018 by the University of Utrecht.⁵⁷ The *CAF-11* and *SyRI* cases, which were ongoing whilst this governmental letter was being written, are not mentioned anywhere. The AI HLEG's *Guidelines for Trustworthy AI*⁵⁸ are mentioned in three footnotes but are not elaborated upon in the main text. The letter only mentions the notion of trust (*vertrouwen*) twice, and does not make use of the concept of *trustworthy AI* as developed by AI HLEG (which bases it in seven key requirements: *human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; environmental and societal well-being; accountability*). Instead the governmental letter proposes a set of its own guidelines, developed in collaboration with experts from administrative agencies. These guidelines aim to increase the understanding, transparency and quality of data analytics used by public bodies; those are added to the letter as a 17-page annex.⁵⁹ The eight areas covered by these

55 Minister for Legal Protection (2019).

56 Minister of the Interior and Kingdom Relations (2019).

57 Vetzo, Gerards and Nehmelman (2018).

58 High-Level Expert Group on AI (2019). The AI HLEG 's guidelines also formed the basis for their *Assessment List for Trustworthy Artificial Intelligence* (2020) following a pilot study with over 350 stakeholders, see Chapter 1 in this volume.

59 Ministerie van Justitie en Veiligheid (2020).

guidelines are: (1) *Awareness of risks* (stigmatisation, bias, hacks, etc.), (2) *Explainability* (in principle public bodies should not use algorithms that are too complex to explain), (3) *Data recognition* (the parameters or training data on which a model is based, should be described and possible discriminatory effects explored) (4) *Auditability* (the models and algorithms that are used for ADM resulting in specific consequences for individual citizens need to be verifiable in case of damages; their use needs to be documented), (5) *Accountability* (public authorities are responsible for decisions made by algorithms, even if these are third-party algorithms or if they cannot be explained in detail), (6) *Validation* (regular testing to see if a model achieves the desired result without creating additional damage), (7) *Verifiability*, (data analytics need to be verifiable - potentially in court), (8) *Provision of information to the public* (information needs to be concise and transparent; variables and threshold values can be communicated as long as this does not bring along the possibility for gaming the system; transparency is more important for ADM that has an high impact/big implications for citizens; transparency is not a holy grail – it can be more important that an independent auditor and the public body itself understand the workings of an algorithm than that an individual citizen understands it).

While there is a clear thematic overlap with the AI HLEG's seven key requirements for trustworthy AI, the list of eight domains requiring safeguards appear to be *locally* grown. It is important to note here that the letter makes an explicit distinction between ethical guidelines and legal regulations for AI used by private companies and those used by public bodies. Regulating the former on a national level seems of little use, given the fact that companies often have activities that transcend national borders. However, with regard to public bodies the government would like to create national regulation or legislation to give the necessary safeguards more legitimacy. The Dutch government considers that the scope of the *General Data Protection Regulation* (2016/269) and *Data Protection Directive 2016/680 for Police and Criminal Justice Authorities* is not extensive enough to provide safeguards with regards

to all the risks involved in the use of AI applications. The government proposes that complementary regulation or legislation might be needed in the following areas: (1) transparency, and (2) quality control. A final focus point is accountability and supervision. The government intends to look at whether there are gaps that are not covered by any supervisory authorities.

To summarise, this governmental letter is based mostly on national expertise and stresses the need for creating (national) legislation and supervision to give safeguards with regards to the use of data analytics by the government. In the Dutch AI strategy's annual review,⁶⁰ the government reports that in 2020 the guidelines are evaluated in terms of effectiveness and feasibility. Furthermore, the question about whether national legislation will be needed will depend on the legislative proposals that will follow from the White Paper on Artificial Intelligence (19 February 2020) in which the European Commission published its digital strategy for 2020-2025. The effectiveness of existing supervisory mechanisms was also considered sufficient. However, the government has suggested exploring to what degree *algorithm impact assessments* could be developed and integrated in already existing *data protection impact assessments*.

The second governmental letter concerns AI, public values and human rights and is written by the Minister of the Interior and Kingdom Relations.⁶¹ In this letter there is no mention of trustworthy AI. Even the word 'trust' is completely absent. Instead the letter stresses the importance of human rights in relation to innovation within the field of AI. The government proposes a human-centered approach to AI that aligns with the European focus on this notion. However, the government also underlines that the current human-centered AI concepts are very abstract and general, and will need substantial concretisation to have an impact on the development of AI technology. The letter expresses the hope that human-centred AI, like other value driven

60 Ministerie van Economische Zaken en Klimaat (2020).

61 Minister of the Interior and Kingdom Relations (2019).

technologies such as wind and solar power, could become an essential export product, and that the Netherlands could become a frontrunner in this field. In the annual review of the Dutch AI strategy the government reports that it launched a Toolbox Ethical Innovation to support (public) organisations.⁶² It also aims to create a human rights impact assessment for AI. In late 2020, the government plans to create a hackathon to create solutions that can combat bias and discrimination in ADM.

6. Conclusions – strengths and weaknesses of trustworthy AI in SAPAI

THE NETHERLANDS HAS been eager to be seen as the vanguard of European AI, especially when it comes to making sure that AI is human-centred. Simultaneously, the national AI strategy does not closely follow the policy documents of the AI HLEG, especially insofar as the concept of trustworthy AI does not play a significant role. Among the strengths of the Dutch AI strategy there are significant financial commitments to AI, and collaborations between the public and private sectors. Similarly, their pragmatic approach and swift promulgation of legislation and supervision appear promising.

The development of the AI strategy, however, has been done by three separate ministries. While one might think that this implies cross-ministerial support for AI, the effect is that there is no overarching vision, the strategy is fragmented, and fundamental values and rights are purposefully kept outside the upbeat glossy SAPAI report. Other areas for concern lie in the Netherlands' recent history of using biased and inaccurate algorithmic tools, as in the *CAF-11* and *SyRI* cases. It therefore appears that the Dutch public sector has had a difficult time in striking the right balance between governmental efficiency and protection of fundamental rights. It is to be hoped that the current strategy will help prevent such cases in the future, and that the focus on human-centred AI will go beyond the creation of a brand and export product.

62 Ministerie van Economische Zaken en Klimaat (2020).

References

- Amsterdam Algoritmeregister Beta. (2020). *Gebruik van algoritmes binnen gemeente Amsterdam*. <https://algoritmeregister.amsterdam.nl/>
- Belastingdienst. (2020). *CAF 11*. <https://over-ons.belastingdienst.nl/dossier-herstel-kinderopvangtoeslag/caf-11/>
- Bronzwaer, S. (2019a). *AI op de agenda, nu de miljarden nog*. NRC. <https://www.nrc.nl/nieuws/2019/10/09/ai-op-de-agenda-nu-de-miljarden-nog-a3976255>
- Bronzwaer, S. (2019b). *Nederland investeert miljarden in kunstmatige intelligentie: Waarom?* NRC Handelsblad. <https://www.nrc.nl/nieuws/2019/10/18/ai-de-beste-voorspeller-heeft-de-meeste-macht-a3977320>
- Chuvpilo, G. (2019). *AI Research Rankings 2019: Insights from NeurIPS and ICML, Leading AI Conferences*. Medium. <https://medium.com/@chuvpilo/ai-research-rankings-2019-insights-from-neurips-and-icml-leading-ai-conferences-ee6953152c1a>
- City of Amsterdam Algorithm Register Beta. (2020). *Holiday rental housing fraud risk*. <https://algoritmeregister.amsterdam.nl/en/holiday-rental-housing-fraud-risk/>
- de Rechtbank. (2020a). *SyRI legislation in breach of European Convention on Human Rights*. <https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx>
- de Rechtbank. (2020b). *ECLI:NL:RBDHA:2020:1878*. <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878>
- Deloitte. (2020). *Nederlandse investeringen in AI blijven achter*. Deloitte Netherlands. <https://www2.deloitte.com/nl/nl/pages/over-deloitte/articles/nederlandse-investeringen-in-ai-blijven-achter.html>
- Deloitte AI Institute & Deloitte Center for Technology. (2020). *Thriving in the era of pervasive AI: Deloitte's State of AI in the Enterprise* (3rd Edition). Deloitte Insights. <https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/about-deloitte/deloitte-cn-dtt-thriving-in-the-era-of-persuasive-ai-en-200819.pdf>
- DutchNews. (2020). *Tax office to be broken up in wake of child benefit scandal*. DutchNews.NL. <https://www.dutchnews.nl/news/2020/01/dutch-tax-office-to-be-broken-up-in-wake-of-child-benefit-scandal/>
- European Commission. (2018). *Communication Coordinated Plan on Artificial Intelligence* (COM(2018) 795 final). https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en
- European Commission. (2020a). *European Innovation Scoreboard 2020*. <https://ec.europa.eu/docsroom/documents/42981>
- European Commission. (2020b). *White Paper On Artificial Intelligence—A European approach to excellence and trust* (COM(2020) 65 final). https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
- Government of the Netherlands. (2019). *Strategic Action Plan for Artificial Intelligence*. The Netherlands. Annex to: Ministry of Economic Affairs and Climate Policy (2019), *Governmental letter on AI strategy*. Parliamentary Papers, House of Representatives 2019-2020, 26 643 and 32 761, no. 640. <https://www.government.nl/binaries/government/documents/reports/2019/10/09/strategic-action-plan-for-artificial-intelligence/Strategic+Action+Plan+for+Artificial+Intelligence.pdf>
- High-Level Expert Group on AI. (2019a). *Ethics guidelines for trustworthy AI* (p. 41). Set up by the European Commission. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

High-Level Expert Group on AI. (2019b). *Policy and investment recommendations for trustworthy Artificial Intelligence* (p. 52). Set up by the European Commission. <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>

High-Level Expert Group on AI. (2020). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*. European Commission. <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

Innovation Center for Artificial Intelligence. (2020). *Homepage*. ICAI. <https://icaai.ai/>

Johnson, K. (2019). How Europe's AI ecosystem could catch up with China and the U.S. *VentureBeat*. <https://venturebeat.com/2019/10/01/how-europes-ai-ecosystem-could-catch-up-with-china-and-the-u-s/>

Johnson, K. (2020). Amsterdam and Helsinki launch algorithm registries to bring transparency to public deployments of AI. *VentureBeat*. <https://venturebeat.com/2020/09/28/amsterdam-and-helsinki-launch-algorithm-registries-to-bring-transparency-to-public-deployments-of-ai/>

Kickstart AI. (2020). *Launching Kickstart AI, a catalyst for artificial intelligence in The Netherlands*. <https://www.kickstartai.nl>

Klein, P. (2019). *Toeslagenaffaire: Topambtenaren gaven illegale fraudejacht ruim baan*. RTL Nieuws. <https://www.rtlnieuws.nl/nieuws/politiek/artikel/4857711/belastingdienst-menno-snel-toeslagen-ministerie-van-financien>

Kleinnijenhuis, J. (2020). *Belastingdienst had een geheime fraudelijst, in strijd met de privacywet*. Trouw. <https://www.trouw.nl/gs-bdda4419>

Minister of the Interior and Kingdom Relations (2019). *Governmental letter on AI, public values and human rights*. Parliamentary Papers, House of Representatives 2019-2020, 26 643 and 32 761, no. 642. <https://zoek.officielebekendmakingen.nl/kst-26643-642.html>

Minister for Legal Protection (2019). *Governmental letter on safeguards against risks associated with data analytics by the government*. Parliamentary Papers, House of Representatives 2019-2020, 26 643 and 32 761, no. 641. <https://zoek.officielebekendmakingen.nl/kst-26643-641.html>

Ministerie van Economische Zaken en Klimaat. (2019). *Strategisch Actieplan voor Artificiële Intelligentie (SAPAI)*. <https://doi.org/10.08/strategisch-actieplan-voor-artificiele-intelligentie>

Ministerie van Economische Zaken en Klimaat. (2020). *Nederlandse Digitaliseringsstrategie 2020* (Annex to Parliamentary Papers, House of Representatives 2019-2020 26 643, no. 709). <https://www.rijksoverheid.nl/documenten/rapporten/2020/06/25/nederlandse-digitaliseringsstrategie-2020>

Ministerie van Financiën. (2020). *Problemen met de kinderopvangtoeslag—Ontwikkelingen Belastingdienst—Rijksoverheid.nl*. Ministerie van Algemene Zaken. <https://www.rijksoverheid.nl/onderwerpen/ontwikkelingen-belastingdienst/problemen-met-de-kinderopvangtoeslag>

Ministerie van Justitie en Veiligheid. (2019). *Brief over waarborgen tegen risico's van data-analyses door de overheid* (Annex to Parliamentary Papers, House of Representatives 2019-2020). Ministerie van Justitie en Veiligheid. <https://zoek.officielebekendmakingen.nl/blg-904102>

Ministry of Economic Affairs and Climate Policy. (2019a). *Strategic Action Plan for Artificial Intelligence (SAPAI)*. <https://www.government.nl/binaries/government/documents/reports/2019/10/09/strategic-action-plan-for-artificial-intelligence/Strategic+Action+Plan+for+Artificial+Intelligence.pdf>

Ministry of Economic Affairs and Climate Policy. (2019b). *Governmental letter on AI strategy*. Parliamentary Papers, House of Representatives 2019-2020, 26 643 and 32 761, no. 640. <https://zoek.officielebekendmakingen.nl/kst-26643-640.html> or <https://www.rijksoverheid.nl/documenten/kamerstukken/2019/10/08/kamerbrief-ai>

Nederlandse AI Coalitie. (2020). Extra financiële ondersteuning voor ambities Nederlandse Digitaliseringsstrategie. <https://nlaic.com/nieuws/extra-financiele-ondersteuning-voor-ambities-nederlandse-digitaliseringsstrategie/>

NL AI Coalition. (2020). About NL AIC. <https://nlaic.com/en/about-nl-aic/>

NL Times. (2019). *Netherlands has EU's most competitive economy: WEF*. <https://nltimes.nl/2019/10/09/netherlands-eus-competitive-economy-wef>

NOS. (2019). *Kabinet wil kunstmatige intelligentie aanjagen met 2 miljard euro*. <https://nos.nl/l/2305235>

NRC. (2020). *Nederland heeft een algoritmewaakhond nodig*. NRC.nl. <https://www.nrc.nl/nieuws/2020/10/14/nederland-heeft-een-algoritmewaakhond-nodig-a4016038>

Oxford Insights and the International Research Development Centre. (2020). *Government AI Readiness Index 2020*. <https://www.oxfordinsights.com/government-ai-readiness-index-2020>

Philips. (2020). *Philips and Ahold Delhaize announce first Kickstart AI joint appointments*. <https://www.philips.com/a-w/about/news/archive/standard/news/articles/2020/20200213-philips-and-ahold-delhaize-announce-first-kickstart-ai-joint-appointments.html>

Raad van State. (2020). *Samenvatting advies wetsvoorstel gegevensverwerking door samenwerkingsverbanden*. <https://www.raadvanstate.nl/@120993/advies-wetsvoorstel-gegevensverwerking/>

Rijksoverheid. (2020). *Maatregelen voor gedupeerden kinderopvangtoeslag—Ontwikkelingen Belastingdienst—Rijksoverheid.nl* [Onderwerp]. Ministerie van Algemene Zaken. <https://www.rijksoverheid.nl/onderwerpen/ontwikkelingen-belastingdienst/maatregelen-voor-gedupeerden-kinderopvangtoeslag>

Rosemain, M., & Rose, M. (2018). France to spend \$1.8 billion on AI to compete with U.S., China. *Reuters*. <https://www.reuters.com/article/us-france-tech-idUSKBN1H51XP>

Stichting Digitale Infrastructuur Nederland. (2020). *Stichting DINL*. Stichting DINL. <https://www.dinl.nl/>

SUWI Act (2013). *Wet van 9 oktober 2013 tot wijziging van de Wet structuur uitvoeringsorganisatie werk en inkomen (SUWI Wet) en enige andere wetten in verband met fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekende zijnde gegevens*, Stb. 2013, 405. <https://zoek.officielebekendmakingen.nl/stb-2013-405.html>.

SUWI Decree (2014). *Besluit van 1 september 2014 tot wijziging van het Besluit SUWI in verband met regels voor fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekend zijnde gegevens met inzet van SyRI*, Stb. 2014, 320. <https://zoek.officielebekendmakingen.nl/stb-2014-320.html>

Taskforce AI (2019), *Algoritmen die werken voor iedereen. Samen bouwen aan onze (digitale) toekomst met Artificiële Intelligentie (Algorithms that work for everybody. Building our (digital) future together with Artificial Intelligence)*. Position paper. Online available at: https://www.vno-ncw.nl/sites/default/files/position_paper_algoritmen_die_werken_voor_iedereen.pdf

Tralac. (2016). *DHL Global Connectedness Index: Globalization surpassed pre-crisis peak, advanced modestly in 2015*. <https://www.tralac.org/news/article/10984-dhl-global-connectedness-indexglobalization-surpassed-pre-crisis-peak-advanced-modestly-in-2015.html>

Tweede Kamer der Staten-Generaal. (2020). *Government assessment of White Paper on Artificial Intelligence* (Annex to Parliamentary Papers, House of Representatives 2019-2020 22112 nr. 2894). <https://zoek.officielebekendmakingen.nl/blg-943554>

van den Hoven van Genderen van Leeuwen, R. (2019). *De “unbearable lightness of being”, het nieuwe Strategisch Actie Plan voor AI*. Nederlandse Vereniging voor AI- en Robotrecht. <https://robotrecht.nvair.nl/2019/de-unbearable-lightness-of-being-het-nieuwe-strategisch-actie-plan-voor-ai/>

van der Starre, B., & den Hollander, D. (2019). Blog: Reactie op het Strategisch Actieplan AI. *Birch Consultants*. <https://www.birch.nl/actualiteiten/blog-reactie-op-het-strategisch-actieplan-ai/>

van Til, G. (2019). *Automating Society: Netherlands*. Algorithm Watch. <https://algorithmwatch.org/en/automating-society-netherlands/>

World Justice Project. (2020). *WJP Rule of Law Index*. <https://worldjusticeproject.org/rule-of-law-index/>

AI policy in Italy

Comprehensive focus on core
infrastructural robustness
and humanistic values

Francesco Cappelletti

Research Fellow at European Liberal Forum;
member of Fondazione Luigi Einaudi.

Summary

THE INTRODUCTION OF an Italian policy for artificial intelligence (AI) has emerged in the broader context of the country's general digitisation strategy, and in the following chapter I also take into account the policy's shared elements with the larger pan-European discussion. Presenting a comparative study, I primarily rely on documents that have been published by Italian Ministers and public institutions, noting that there is a general coherence with those published by the European Commission. The Italian Ministry for Technological Innovation and Digitalisation (MID), in the "Piano Nazionale Innovazione 2025", explicitly places the Italian AI strategy in the broader context of digitalisation of the whole country. The document clearly depicts an idea of "artificial intelligence [that aims

to be] socially, culturally and democratically sustainable”, at the heart of the digitalisation process. The main document guiding the Italian approach to AI, the *Italian Strategy for Artificial Intelligence*, published by the Ministry for Economic Development (MISE) in July 2020, takes into consideration many aspects related to the positive approach of this new technology “with a clear anthropocentric imprint and oriented towards sustainable development”. There is, therefore, a clear discursive match with the principles listed in the European Commission’s High-Level Expert Group’s documents concerning the *Ethics Guidelines* and principles for a “trustworthy AI”. Additional similarities can be identified with the Commission’s later recommendations. Finally, it should be taken into account that the Italian AI strategy aims at a project involving different institutions and ministries. The digitalisation process in Italy is thought to purposefully re-innovate the whole public administration preparing the country for the emergence of new technologies. Research and partnerships between actors (both public and private), involved in these developments, will play a fundamental role in creating an integrated ecosystem for the advancement of the skills required to build the new digital infrastructure. Therefore, other important legislative and policy documents are taken into consideration in the same way as those just mentioned.

1. Introduction

THIS CHAPTER AIMS to provide a general perspective on the main developments and ongoing projects concerning AI policy in Italy. My analysis is intended to highlight the most recent developments, presented by the Ministry of Economic Development (MISE), regarding the official AI strategy for Italy,¹ and the related digitalisation plan for the public administration, proposed by the Ministry of Technological Innovation and Digitalisation (MID), which tends to act as a link for all the various

1 MISE (2020), the 82 proposals for an AI strategy contained into this document were drafted in 2019, to be submitted for consultation and approved in February 2020, and published, with further updates, in July 2020.

actors involved in this ambitious project.² In the same way, the most recent regulatory developments will be examined to understand how government policies are currently designed and implemented.

Some factors are crucial to understand the general state-of-the-art Italian digitalisation and AI policy. First, the European Index on Digital Society (DESI), shows a general lack of digitalisation in Italy compared to many other European countries – especially in terms of network access, business digitisation and connectivity; factors where Italy stands below the European average.³ Second, the internal composition of the productive structure of the country is very dependent on small and medium-sized enterprises (SMEs) – constituting around 90 per cent of total productivity.⁴ Although the available data will have to be reassessed in the light of the sudden digitisation experienced by Italian SMEs during the 2020 pandemic, indicators related to online trade and use of digital marketing tools place Italian SMEs at the bottom of the list of European countries. Third, what is relevant to any discussion on AI is the use of “big data.” Also in this case, a sense of backwardness emerges from the comparative data, when considering the current uses of such technologies in Italy.⁵

2 MID (2020a, 2020b).

3 DESI (2020) is a composite index that summarises relevant indicators on Europe's digital performance and tracks the evolution of EU Member States.

4 In theory, Italian companies that contribute with a turnover of less than 50 million EUR per year. However, it should be remembered that small Italian companies are part of complex value chains that go beyond the domestic market. (Info Data, 2019).

5 While the data market in Italy is relatively small compared to the potential of the domestic product, and only 12 percent of Italian companies have commenced projects that include the use of AI, it should be remembered what has been said above about the composition of the Italian productive network, in which small companies make up a large part of the total. It is therefore not correct to speak of lack of investment in data as a sign of backwardness, since many of the country's strategic companies (both in the technological sector and in manufacturing – robotics, automotive, automation) are at the forefront in terms of technological advancement.

2. The digitalisation process in Italy

BEFORE INTRODUCING THE discussion on Artificial Intelligence, it is useful to briefly describe the current situation in Italy, regarding recent regulatory developments affecting new technologies and digitalisation processes. One of the key elements of Italy's digital development is its Public Administration sector, intending to offer citizens full access to digital services. Coordination is provided by the Agency for Digital Italy (AGID).⁶ Specifically, the AGID operates through the structuring of three-year plans, approved by the government, which has as its general objective the digitisation of key national infrastructure. The most recent plan, the third of its kind ratified by the Prime Minister in 2020,⁷ with the aim of “developing a digital society [...] for citizens and businesses,” and promoting “sustainable digitalisation”, contributing to the diffusion of new technologies “in the productive Italian network”.⁸ Also, AGID will oversee the management of Digital Identity for Citizens and simplified payment systems for public administration. It should be noted that in 2018 the agency issued a White Paper on the topic of “AI at the service of the Citizen”, containing some guidelines included in the most recent strategy.⁹ In particular, in the White Paper, it was recommended to identify the minimum *safety* measures for the implementation of this technology, as well as to evaluate the usability of data from different administrations, until the development of the necessary *skills*.¹⁰

While Italy seems to be catching up with the rest of the European countries in terms of new technology development, only recently the government has developed a solid internal structure to address the

6 The Agenzia per l'Italia Digitale (AGID) is the governmental agency that coordinates the implementation of the Italian Digital Agenda and promotes the spread of the use of ICT technologies by promoting innovation and the economic, social and cultural growth of the country (www.agid.gov.it). The AGID is dependent on the Ministry for Technological Innovation and Digitalisation.

7 The Italian Prime Minister and the Minister for Technological Innovation and Digitalisation (2020).

8 MID (2020b).

9 AGID (2018).

10 Ibid., pp. 75–77.

issue of digitalisation, specifically through the creation of dedicated agencies. One of them, the Department for Digital Transformation,¹¹ is the main actor to support the role of the Prime Minister and to implement the actions necessary to realise the agenda points of the three-year AGID plan.

However, the digitalisation process in Italy is not limited to a single action. There are several initiatives aimed at digitising markets and involving businesses and citizens. The most ambitious, promoted through the MID, and addressing all fields of technological development, is called “Italia 2025.”¹² The Department for Digital Transformation has created a development plan within the project that puts into practice the digitalisation proposals already outlined in previous years and supported by the European institutions. The platforms under construction inside this project aim at a 360° digitalisation for both the public administration and for individuals, to manage all available services digitally.

In this framework, a crucial point for the creation of a highly digitised country is the creation of a tailored and robust infrastructure. For this purpose, a Public Connectivity System (SPC)¹³ is outlined, which aims at *interoperability* of infrastructures involving both the Public Administration, and other (private) actors in favour of the implementation of the system as a whole. This interoperability is at the heart of AGID’s three-year IT plan. The AGID is responsible for governance and

11 This institution, within the MID, was established by the Decree of the Prime Minister (DPCM of 19 June 2019) and made effective by the Secretary General of the Presidency of the Council of Ministers (Decree of 24 July 2019). For more information, see <https://innovazione.gov.it/it/chi-siamo/dipartimento/>.

12 MID (2020a).

13 The Sistema Pubblico di Connettività (SPC) defines the preferred ways in which public administrations’ information systems must be interoperable with each other. The company that manages the entire system is CONSID, of which the Ministry of Economy and Finance is the sole shareholder. CONSID has entrusted the management of the service through a public procurement contract to a temporary joint venture (composed by Fastweb S.p.A., Sistemi Informativi S.r.l. and Finmeccanica S.p.A.) which, since 2016, is in charge of the provision of interconnection, interoperability and governance services under the aegis of the AGID (and the MID for extension). Further information on AGID website: <https://www.agid.gov.it/it/infrastrutture/sistema-pubblico-connettivita>.

oversees critical aspects of the system (software versioning, traceability of requests, adequacy of the service, coordinated API documentation).¹⁴ At the core lies the assumption that administrations and stakeholders ought to adhere to established international and European standards.¹⁵

These innovation plans follow recent regulatory developments at both European and international level. It emerges clearly from the official statements that the country's technological development must retain the “sustainability” of projects and make them “ethical,” in line with the United Nations' Sustainable Development Goals (SDGs).¹⁶

2.1. An *ad hoc* structure for Italy's digitalisation and AI

Like any other new technology, the development of AI requires new legislation. This concept seems to be even more relevant considering the improvement of AI. There is a general challenge for the development, implementation, and uses of technology to be implemented through ethical standards that are aligned with the values of the society in which we live.¹⁷ The European Commission's recommendations seem to remain vague regarding the implementation of domestic regulations for the introduction of AI.¹⁸ Hence, the Italian strategy for AI anticipates some possible solutions, in line with a shared approach at the international level, while still retaining the objective advantages in terms of the shared reliability of these new technologies. Legislation, nevertheless, cannot be reduced to mere mathematics,¹⁹ it must follow a specific logic: The regulator should fully understand in advance the possible consequences and impacts of a new sort of technology, especially when – as in the case of AI – it affects daily life in ways never seen before. A thriving legislative system can only be achieved through strategic planning of AI policies, taking into account medium and long-term effects.

14 The new interoperability framework is adopted by the agency, as outlined in AGID (2020).

15 Especially taking into account the ISA² - European Interoperability Framework (on the basis of COM 134/2017), see European Commission (2017).

16 MID (2020a), p. 3.

17 IEEE (2019).

18 Although some key points in the Commission's guidelines refer to issues covered by current legislation, see Larsson (2020).

19 Hagendorff (2020).

The focus must therefore be on innovation, which must be favoured from a regulatory point of view.

Innovation, in the pursuit of creative solutions, is seen as an essential part of the process of development of an AI that must be “sustainable” and at the same time “understandable” by the citizen.²⁰ The MISE documents explicitly refer to the use of technology that is Human-Centered. This is particularly relevant in relation to the European Commission’s High-Level Group of Experts (AI HLEG)²¹ and aligns the Italian strategy with the EU recommendations. This leads to some aspects being considered, primarily concerning the awareness and education of the citizens, and also with the aim of being prepared to make conscious decisions through democratic processes.

One of the central remaining points of the national strategy is the issue of governance. The MISE document states how it will be necessary to think of an institutional structure to supervise the process and implementation of digital policies, especially regarding AI. The concepts listed above can be found both in the Italian strategy for AI and in the European guidelines. The rationale is to allow the regulator to have a constant *evaluation*, not only of the actual implementation of the necessary legal improvements but also of the assessment of the impact and possible risks. The supervisor of these processes should also take into account the *partnership* between public and private sectors, involving research on and experimentation in the decision processes, involving both large and small businesses with the aim of better directing both policies and investments in future projects within the sector.

In general terms, the need for *innovation* mentioned above requires resources, decisions and political direction, but above all a certain capacity to speed up decision-making processes. Still speaking in very general terms, we can identify factors that could disadvantage the process of new technologies in relation to legislative power, the first

²⁰ Ibid. MISE (2020), p. 7.

²¹ See European Commission (2018b, 2019, 2020), AI HLEG (2019a, 2019b, 2020a, 2020b).

concerning the political cycle, the second the readiness to respond to needs, the third linked to the speed of technology development. The political cycle, whose agenda may change unexpectedly, can take on unforeseen configurations, with a change in political direction, which could lead to the interruption of long-term projects. Moreover, for the legislator, it will be essential to respond adequately to requests from the world of research and industry. To this end, the Italian strategy foresees a system of constant dialogue with the parties involved in AI processes. Finally, the speed of technology development is certainly not comparable to that of the approval of a decree in the legislative chambers. This partly explains the need, elaborated in the Italian Strategy, for the creation of a body under the Council of Ministers, i.e. making the institutions at the centre of the (ministerial) functions resilient, modelling the agenda in accordance with requests or feedback from the sector of interest.

An example of smart governance is a recent decree introducing “urgent” measures for digital innovation, which could prove essential for all digitisation processes. The new features contained in the Law Decree 76/2020 approved by the Italian Parliament not only simplify the methods of presenting new projects having “[...] emerging technologies and initiatives with high technological value [...]”, but also open doors to what is called “authorised *experimentation*” in the field of innovative research. This involves companies, universities and research centres, research bodies and start-ups.²² The decree that brings these simplifications is unique in that it enables the digital sector to be divided from other sectors that use public procurement, and thus makes the entire process smoother as part of a long-term strategy – which potentially means being more attractive for investments within the sector.²³

22 The President of the Republic of Italy (2020).

23 The decree was approved just after the so-called *phase 2* of the COVID-19 lockdown in Italy, which had at the time led to a general negative drop in GDP. In this sense, the “urgency” of the amendment must be understood as an attempt to face the crisis by encouraging digitalisation processes in order to boost the related market, fostering the introduction of new technologies in the whole productive sector.

We can thus understand *ad hoc* not only as a call for a greater degree of regulatory experimentation, but as an admission that regulation is historically contingent. It is clear that in the broader context of digitisation, future legislators will need to take into account historically unique issues, but which may become cornerstones for introducing this technology into the country with *corrective* effects on issues that have been bureaucratically difficult to deal with up until then. AI may in this sense become the key to a future *smart government*, which will help the regulatory steps be in line with the prerogatives agreed at European level, but which, at the same time, will allow the country to enhance its areas of excellence – first and foremost, its ethos of *made in Italy*.

3. Artificial Intelligence in Italy

AS ALREADY DESCRIBED, the strategy for the introduction of AI in Italy follows the recent developments carried out by the MISE. The strategy emerges through a series of proposals and pragmatic, flexible points – therefore in progress – which aims at bringing AI into the country following a gradual path. The MISE policy document identifies an AI ecosystem in Italy based on three pillars, i) research and technological transfer, ii) production and iii) adoption. *Research* is understood as wide-ranging and includes the entire university sector and public and private research centres. *Production* means both software development and components, as well as infrastructure and services. The usage (or *adoption*) is referring to the part of the users of the technology. This includes both public administration (in the broader project of digitalisation of the country) and industry and manufacturing.²⁴

Some main strategic parameters can be identified by comparing the different digitisation projects and the strategy for AI. The key points of the strategy as a whole are:

²⁴ MISE (2020), p. 28-30.

- ▶ Encouraging high-level research by attracting talent and enhancing excellence through the development of digital innovation hubs, fostering collaboration between universities and industry;²⁵
- ▶ Support innovative policies for SMEs and start-ups with high technological value and to favour the national industry of AI, also supporting the decentralisation of projects (in regions, provinces, municipalities), generating a growing mass of intellectual resources;
- ▶ Strengthen the use of artificial intelligence-based technologies in sectors of excellence (robotics, automotive, industrial automation) by promoting the conditions to attract foreign investment;
- ▶ Promote the development of embedded AI solutions, advanced cloud and data storage systems, use of internet-of-things (IoT) devices through the growth of the relevant industry sector

All the strategic recommendations for the implementation of an AI ecosystem must follow mandatory guidelines in terms of *ethics* and *sustainability*. In the same way, technology must be seen to have a positive impact on society, aiming at improving the quality of human life (with particular reference to local and territorial challenges). Specific consideration within the Italian strategy for AI is given to *accessibility* and *social inclusion*. Notably, it is proposed to focus research for people with disabilities, implementing AI-based IoT technologies, tools and devices for the inclusion of these people, with particular reference to local and territorial policies. Another interesting possibility of application could be the use of technology for the inclusion of refugees and asylum seekers in society, through intensification in the use of database platforms for an increasingly optimised management of hospitality – and integration.²⁶

3.1. Further implementation for Italy

The planning documents for AI in Italy identify this technology not as a separate element in the field of new technologies, but as an *ecosystem* of

25 Italy has centres of excellence in the technological field such as the National Laboratory of Artificial Intelligence and Intelligent Systems (AIIS, which operates within the Consorzio Interuniversitario Nazionale per l'Informatica; CINI), and infrastructures such as CINECA, GARR and specialised laboratories such as INFN.

26 MISE (2020), part 6.2.

technologies in which AI is implemented. Given the existence of this vision, the implementation of the Italian AI strategy will be dependent on an inter-ministerial governance body (“cabina di regia”) with the Presidency of the Council of Ministers as its final point of reference.²⁷ Behind this logic lies an understanding of AI as theoretically variegated, touching on different areas (from robotics to industry to cybersecurity) in which technology is used and is therefore not the only factor. Consequently, since it is developed in different areas, there is a need for inter-ministerial control, to oversee the whole process of the digital transformation of the country.

As a further development of this vision, the strategy for AI in Italy foresees the creation of an Italian Institute for Artificial Intelligence (I3A).²⁸ This new structure will become a reference point for the development of AI in Italy as part of the strategy itself. The programmatic points of this structure are listed in the design:²⁹

- ▶ developing *top-quality research*;
- ▶ attracting *funding* from industry and the European community;
- ▶ creating a *network for excellence* by collaborating with existing realities in Italy;
- ▶ developing a *technology transfer system*

The MISE’s strategic plan defines the I3A as responsibilities on the application of AI in robotics and industry, healthcare, sustainability/energy, agri-food, transport, aerospace and defence, public administration, and culture / digital humanities.³⁰

3.2. Cutting-edge technology and AI

As mentioned above, a Public Connectivity System will be at the basis of the country’s digitalisation processes and will foster a continuous

27 MISE (2020), pp. 78-80.

28 Although the establishment of such a structure is “desirable” according to the MISE Strategy, the creation of this structure is currently being defined, and its operational headquarter will probably be in Turin.

29 MISE (2020), pp. 80-81.

30 MISE (2020), p. 37.

exchange of know-how within the institutions. A fundamental element in the development of AI is the capacity for high performance of the computing grid. One of the tasks for the I3A project will be to ensure the creation of a computing architecture at the domestic level. This should be done through close collaboration with European computing infrastructures.³¹ The optimal strategy for Italy, in its technological modernisation, seems to be based on a mixed system of High-Performance Computing (HPC) and edge technology. Since machine learning – the core technology of AI – requires the processing of large amounts of data, a strategic distribution of computing processes (as in the case of so-called “edge computing”) could lead not only to greater precision but could also ensure greater energy efficiency to the system. Certainly “the computing capacity must be commensurate with the ambitions [of the strategy]”.³²

The cutting edge of technology finds a European dimension in the Euro HPC Joint Undertaking initiative³³ which aims to create a series of highly efficient and advanced technology computing infrastructures. The project, carried out jointly by the European Commission, several countries and private actors, involves the creation of supercomputers using so-called exascale technology³⁴ in Bulgaria, Czech Republic, Italy, Luxembourg, Portugal, Slovenia and Spain.³⁵

The Italian supercomputing infrastructure will be provided by CINECA,³⁶ a consortium including the Ministry of Education, about 70 universities,

31 Specifically, reference is made to EOSC, Euro HPC and the project of a European infrastructure of supercomputing systems (European Data Infrastructure), see European Commission (2018a).

32 MISE (2020).

33 MISE (2020), p. 83.

34 *Exascale supercomputers* represent a significant milestone for computer engineering. Specifically, reference is made to computers capable of generating calculation operations up to 1 exa-FLOPS, corresponding to a quintillion (10¹⁸) of floating-point operations per second, Gagliardi et al. (2019).

35 The European Joint Undertaking on High-Performance Computing.

36 CINECA is a partner in several pan-European high-performance computing projects.

INFN, and SISSA.³⁷ Leonardo (as it will be named), will be constructed in Bologna by Atos in collaboration with Nvidia and will be part of the European HPC network. It will be used within the Joint HPC initiative for pharmaceutical research, space exploration and weather modelling. It is also important to stress that part of this immense computing capability (which places it among the most powerful supercomputers in the world) will also be available to Italian companies and research. CINECA is committed to promoting the country's digitalisation processes and the transmission of HPC technologies also to the broader industry.³⁸ In a similar way, this provides institutional support for implementation of AI technology within the SMEs.³⁹

The technological impact of high-performance computing initiatives can prove to be fundamental to stimulate the use of AI within the Italian productive sector, potentially leading to a spill-over effect, in which a competitive race towards equitable AI implementations could occur. It is not to be ruled out that, over a few years, a possible spread of this technology to the SME sector will also occur (which, as we have seen, represent the majority of the total number of Italian companies), allowing for better design and production of services, as well as a significant enlargement within the European market.

3.3. Shared and reliable projects:

AI between technology and ethics

The *ethical* question with regards to AI is a cornerstone of the whole discussion, both at European and national level. There are many concerns, especially on behalf of the end-user, about the possible spin-offs of this technology. A lack of clarity is indeed perceived regarding the use of data, and therefore privacy – especially when it comes to the

37 The project for the Leonardo system was presented by CINECA in agreement with the Italian Ministry of Education, University and Research, the National Institute of Nuclear Physics (INFN), and the International School of Advanced Studies (SISSA).

38 Collaboration projects will involve most advanced Italian companies, such as Eni and Bi-Rex, Ferrari, Piaggio Aero, Fiat, BMW Oracle, Luna Rossa, Bonfiglioli Riduttori, see Balocchi (2020).

39 The FF4EuroHPC project, of which CINECA is a partner, is co-financed by Joint HPC and the European Commission.

automation of intelligent systems. This is due to the presence of a market that tends to be monopolistic in which a few large technology companies, using ubiquitous technologies dealing with massive amounts of data, control the privacy of the users (so-called *data hoarding*).

To ensure *reliable* implementation of AI, the MISE Strategic Plan – both in its first version in 2019 and in the updated 2020 version – proves to be consistent with the principles for a *trustworthy AI*, as identified by the AI HLEG.⁴⁰ The MISE Strategy defines the concept of trustworthy AI in parallel with some of the fundamental ethical principles in the field, and could therefore be said to be implemented by the strategy.⁴¹ Moreover, the group of experts of the Italian Ministry of Economic Development stresses that the EU’s targets for trustworthy AI are to be aimed for, but should not be seen as necessary conditions to enter the European AI market at the moment.⁴² The applicability of these principles, therefore, although desirable, is only part of the complex system of guarantees necessary to implement this technology in Italy.

This does not mean that the AI HLEG’s considerations do not underpin the strategy. In the AI HLEG’s Policy and Investment recommendations document,⁴³ there are some points in common with the Italian strategy. Simultaneously, the Ministry of Economic Development tries to pursue a broader vision, taking into account the different needs within the Italian productive sector. In general, the focus is on creation of product certifications that are in line with European recommendations – by various means – but, as described below, aiming to shape the whole production chain, sharing responsibility between all actors involved.

40 Ibid. MISE (2020), art. 2.2.1.

41 The strategy does not examine the ethical issue in detail from a single perspective, and takes into account the key principles from various statements, including the “Asilomar Principles on AI”, the IEEE document on “Ethically Aligned Design”, and the EGE Group’s “Statement on AI, Robotics and Autonomous Systems”.

42 MISE (2020), p. 23.

43 AI HLEG (2019b).

The MISE strategy explains that, at present, the discussion about AI's reliability at national level focuses on the ethics of this technological tool throughout the whole process of development, conception and implementation of the system (*ethics by design*). This is, however, described as “reductive”, as the safety and robustness requirements of the technology must also be taken into account, as indicated by the Commission's AI HLEG.⁴⁴ Therefore, the discussion must concern the technical engineering areas, committing the actors to submit to minimum requirements in terms of ethical compliance, legal standards and technical requirements. A suggested solution could be to create an impact assessment and empowerment tool involving all actors in the AI technology development chain.⁴⁵ This would ensure accountability of all actors engaged as they are called to provide an assessment throughout the entire life cycle of the system (not only *ex-ante*), including also operators and users.⁴⁶

Furthermore, the recommendations contained in the MISE strategy offer a detailed imprint that tends to align national and European standards and, at the same time, gives some propositions on how Italy can play a central role in future ethical discussions regarding AI.⁴⁷

THE MAIN POINTS

- ▶ Adopting the *Trustworthy AI Impact Assessment* is a necessary step because it allows actors to contribute to risk analysis, according to the type of technology implemented (and the risk generated);⁴⁸
- ▶ Strengthen the tools for B2B contracts that can promote risk-sharing along the supply chains;
- ▶ Make private certifications in line with the principles for a *trustworthy AI*;
- ▶ Protect consumers (advanced dispute resolution mechanisms, guarantee privacy, detect misleading advertising);

44 MISE (2020), p. 54.

45 This idea takes shape within the strategy from a possible implementation of the Data Protection Impact Assessment (DPIA) and the privacy impact assessment of the GDPR, which could be adapted to AI policy, see MISE (2020), part 5.1.

46 The evaluation grid referred to is the ALTAI, created by the AI HLEG, which includes commitments in terms of quality, updating and variety of the data used, see AI HLEG (2020a).

47 MISE (2020), pp. 93–94.

48 AI HLEG (2020a).

- ▶ Keeping the data used for machine learning and training within the Italian/European territory;
- ▶ Comply with any legislative shortage

A fundamental role, therefore, will be performed by the public sector, which will have to promote this technological development, but at the same time guarantee the reliability of the entire ecosystem –especially the institutional one– providing the necessary conditions, including legislative ones, to ensure safety from the service provider, the service developer to the consumer. The first step, as mentioned above, could be the creation of the strategic coordination body (“cabina di regia”) mentioned above to oversee the quality of regulations and research. With a medium- and long-term objective in mind, this body should also be able to contribute to the monitoring of Italy’s commitments at the European level, while ensuring the sustainable development of the technology.

Another factor to consider in order to conclude this section is that of the consumers, more specifically: the end-users and the degree with which the strategy intends to develop the awareness regarding the use of AI-based applications.

First, as for the discussion on digitalisation processes and strategic experimentation, the current state-of-the-art domestic law must be taken into account. Within the European regulation, specific rules or directives on civil liability related to the use of machinery or applications using artificial intelligence are not identifiable.⁴⁹ An article in the Italian Civil Code concerns dangerous activities that refers to non-contractual activities (article 2050 of the Italian Civil Code does not exclude from the list of hazardous activities the use of systems that relate to human beings but does not specify anything about the use of AI and robotics),⁵⁰ although there are no direct references to the use of

49 The European Parliament has proposed a compulsory insurance scheme: European Parliament (2017).

50 Reference is made to »dangerous activities«, therefore an intrinsically a risk-bearing activity. This clarifies that those who carry out this type of activity must be considered – in general terms – more exposed and therefore (also in the case of AI) responsible.

AI at present. On the other hand, as far as contractual practices are concerned, there are different internal regulations, but the strategy recommends action at European level for new directives in this area.⁵¹

Second, within the recommendations of the MISE for the protection of consumers and users, we can identify i) the introduction of obligations to share specific datasets, ii) the protection against damage resulting from automated decisions, iii) the protection of the individual from misleading advertising or covert online techniques. The necessary regulations should, therefore, be implemented.⁵² The MISE strategy also lists projects to raise public awareness of AI. This idea, which takes into account other successful initiatives internationally, will depend very much on the degree of interaction that users have with the digital world. Finally, making the digitalisation process inclusive is the basis of the MID *Italia 2025* strategy, which means also creating new skills and reskilling not only for workers but also for users of digital services.⁵³

Finally, the need to adopt a technology should be followed by a process to educate citizens, favouring *computational thinking*, which ultimately leads to the technology being used to enable the individual to fulfil himself or herself in society and his or her tasks. This particular aspect sees Italy in a backward position, especially regarding the issue of education. Therefore, the strategy takes into account the need for concrete action to create *awareness* both in educational processes promoting initiatives in this direction.⁵⁴

3.3.1. AI and the role of Big Data

Although the strategy stipulates that the state must cover a fundamental role, favouring digital and technological processes with adequate legislative responsiveness, it is the *Italy 2025* strategy (of AGID and MID) that clarifies the role that this technology could have for the state

51 MISE (2020), p. 58.

52 MISE (2020), pp. 46-48.

53 MID (2020a).

54 MISE (2020), pp. 90-91.

in detail. Although this is a project in its starting phase, it is clear that the intention is to use “big data” wisely, which hopefully leads to “guiding public decision-makers towards conscious choices”,⁵⁵ in the repetitive processes of public administration, with particular reference to the world of justice. Given the slowness of legal proceedings for which Italian jurisprudence is famous, it is plausible that an implementation – any implementation – in these administrative processes can improve the state of affairs.

One of the main obstacles in absolute terms to this progress concerns the data economy in Italy. As recent studies have shown, most advanced Italian companies need skills (in terms of personnel) within the company. While training (both in-house and university) could be the solution, at the same time companies are reluctant to deal with the data market and data sharing for security reasons.⁵⁶ It is plausible that the choices, not only strategic but also political, that have been described so far can promote the further development of this market, and those Italian companies will benefit from the opportunities offered by the data economy. This should also be seen in the broader context of the possible advantage that the widespread use of HPC technologies could have in the production structure.⁵⁷

3.3.2. The Italian *RenAIssance*

The Italian strategy speaks of a real *renaissance* for the economy that could be made possible through the use of AI. This so-called *RenAIssance*, proposed by the government, foresees a data-driven innovation, in which the institutions and the national government have a pivotal role to play. As was already discussed, the possibilities of offering better decision-making processes, through *evidence-based policymaking*, can lead to smarter regulation. One more step can be acknowledged in the MISE strategic plan, not only related to new regulatory actions for

55 MID (2020a), p. 18.

56 A study AI of applications to the Italian market is given by the AI Observatory of the Polytechnic of Milan, Osservatori.net (2019).

57 See section “Cutting-edge technology and AI” in this document.

technologies and industrialisation but in general looking at the medium and long term, to address rationalised policies aimed at achieving absolute benefits. The introduction of these decision-making methodologies could pass through controlled environments (*regulatory sandboxes*) and would allow operators from different sectors such as universities, experts and companies to interact. This makes it possible to verify *ex-ante* the sustainability and applicability of a new regulatory project.

Moreover, further facets of the strategy can be traced in the document of the MID, *Italy 2025*, where it is proposed that an agency is to be created, “Alliance for sustainable Artificial Intelligence”,⁵⁸ which is thought to act as a supervisor on an ethical-legal set of rules, providing feedback on the ethical sustainability of technological implementations. Although this project departs from the MISE strategic programme, it is essential to stress the multidisciplinary nature of the subject.

4. Conclusions

THE PROGRAMMES FOR the digitisation of the country and the strategy for the implementation of AI in Italy can be defined as comprehensive. Although many technical issues have not been addressed in this short chapter, in general, the Italian strategy for AI appears to be based on solid foundations. It notes how many implementations are still needed.

One of the key remits of the policy is its focus on the implementation of AI for optimising management of administrative practices – and therefore also public administration – and to provide, also, specific responses to local needs, both regional and provincial. This last point seems fundamental when we consider the economic disparity between the Italian north and south and the particular characteristics of the various regions. The success of these decentralised projects will depend on the quality of the future infrastructure, which must take into account the geographical particularities of some areas.

⁵⁸ MID (2020a), p. 25.

From a market point of view, it is clear that Italy and its productive apparatus will benefit from a strategic application of AI in the productive sectors and industry, also boosting the growth of know-how, and high-level research within universities and research centres.

The sensitive issues related to the use of this technology are also addressed comprehensively, with particular reference to the problem of reliability and inclusiveness. Also, as regards to *sustainability*, it is defined as a broad concept. It will depend on the degree of development of the relevant technologies within AI in general and the subsequent digital ecosystem, but always taking into account the European reference standards.

Comparing the Italian strategy with what is described in the European Commission's documents, a clear alignment with the basic principles of the European recommendations can be identified. At the same time, as already described, an attempt is made to structure the approach to AI according to the specific needs of the country. Finally, there is an evident readiness on the part of Italian institutions to "be promoters" of this development,⁵⁹ through the sharing of best practices and participation in the European debate. The documents of the AI HLEG have been therefore clearly inspiring in outlining the Italian strategy, especially concerning the issue of the *trustworthiness* of technology and the *anthropocentric* approach it must have.

A singular feature of the Italian strategy is that it is firmly based on its political direction and commitment. At the same time, a strong focus is given to the "listening policy," especially for having feedback from the production sector. The balance between the political agenda of the future (how much future governments will want to follow this path) and the effectiveness of the productive sector of the country to be incisive on the forthcoming legislative steps, will determine the success of a real *smart policy* concerning the national AI implementation.

59 Both in MISE (2020) and in MID (2020a).

References

AGID (2018). “Libro Bianco sull’Intelligenza Artificiale al servizio del cittadino”. White Paper, version 1.0, March 2018. Available at <https://ia.italia.it/assets/librobianco.pdf>, accessed October 20, 2020.

AGID (2020). “Determinazione 406/2020: Adozione della Circolare recante le linee di indirizzo sulla interoperabilità tecnica”. Circular, Rome, September 9. Available at https://trasparenza.agid.gov.it/moduli/downloadFile.php?file=oggetto_allegati/202581145100O__ODT+DG+n.+406+-9+sett+2020+-+Circolare+1-2020+Linea+di+indirizzo+interoperabilit%Eo+tecnica.pdf, accessed October 20, 2020.

AI HLEG (2019a). Ethics Guidelines for Trustworthy AI. Brussels: European Commission’s High-Level Expert Group on AI, April 8. Available at <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, accessed October 16, 2020.

AI HLEG (2019b). Policy and investment recommendations for trustworthy Artificial Intelligence. Brussels: European Commission’s High-Level Expert Group on AI. 26 June. Available at <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>,

AI HLEG (2020a). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Brussels: European Commission’s High-Level Expert Group on AI. July. <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>,

AI HLEG (2020b). Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence. Brussels: European Commission’s High-Level Expert Group on AI, Deliverable 4. July. Available at <https://futurium.ec.europa.eu/en/european-ai-alliance/document/ai-hleg-sectoral-considerations-policy-and-investment-recommendations-trustworthy-ai>,

Balocchi, Andrea (2020). “Cineca, dove la forza del supercomputer incontra l’industria”, *Industria Italiana*, May 8. Available at <https://www.industriaitaliana.it/cineca-supercomputer-industria-supercalcolo-hpc-digitalizzazione/>, accessed October 21, 2020.

DESI (2020). The Digital Economy and Society Index 2020. European Commission. Available at <https://ec.europa.eu/digital-single-market/en/desi>, accessed October 20, 2020.

European Commission (2017). New European Interoperability Framework: Promoting seamless services and data flows for European public administrations. Luxembourg: Publications Office of the European Union. Available at https://ec.europa.eu/isaz/sites/isaz/files/eif_brochure_final.pdf, accessed October 20, 2020.

European Commission (2018a). »Proposal for a Council Regulation on establishing the European High Performance Computing Joint Undertaking«. COM(2018) 8 final. 2018/0003 (NLE). January 11. Available at https://eur-lex.europa.eu/resource.html?uri=cellar:c48188c9-f6bb-11e7-b8f5-01aa75ed71a1.0001.02/DOC_1&format=PDF,

European Commission (2018b). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Coordinated Plan on Artificial Intelligence (COM(2018) 795 final). Brussels, December 7. Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:795:FIN>, accessed 28th September 2020

European Commission (2019). “A definition of Artificial Intelligence: main capabilities and scientific disciplines”. Available at <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>, accessed 16th October 2020.

European Commission (2020). »On Artificial Intelligence - A European approach to excellence and trust«. White paper, COM(2020) 65 final. Brussels, February 19. Available at https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf,

European Parliament (2017). "Civil Law Rules on Robotics". Strasbourg: European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). Available at https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html#title1, accessed October 21, 2020.

Gagliardi, Fabrizio; Miquel Moreto; Mauro Olivieri; Mateo Valero (2019). "The international race towards Exascale in Europe". CCF Transactions on High Performance Computing: 1, 3-13.

Hagendorff, Thilo (2020). "The Ethics of AI Ethics: An Evaluation of Guidelines". Minds and Machines, 30: 99-120.

IEEE (2019). A Call to Action for Businesses Using AI: Ethically Aligned Design for Business. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. March. Available at https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead/ead-for-business.pdf?utm_medium=PR&utm_source=Web&utm_campaign=EAD1e&utm_content=geias&utm_term=undefined, accessed October 21, 2020.

Info Data (2019). "Pmi, quanto conta in Italia il 92% delle aziende attive sul territorio?". Il Sole 24 Ore, Info Data section, July 10. Available at <https://www.infodata.ilssole24ore.com/2019/07/10/40229/>, accessed October 20, 2020.

Larsson, Stefan (2020). "On the Governance of Artificial Intelligence through Ethics Guidelines". Asian Journal of Law and Society, 1-15. Peer-reviewed postprint, 18 May .

MID (2020a). "Piano nazionale innovazione 2025: Release stabile". Italian Ministry for Technological Innovation and Digitalisation. Rome, February 13. Available at <https://docs.italia.it/media/pdf/piano-nazionale-innovazione-2025-docs/stabile/piano-nazionale-innovazione-2025-docs.pdf>, accessed October 20, 2020.

MID (2020b). "Piano Triennale per l'informatica nella Pubblica Amministrazione". Italian Ministry for Technological Innovation and Digitalisation. Rome, July. Available at https://www.agid.gov.it/sites/default/files/repository_files/piano_triennale_per_linformatica_nella_pa_2020_2022.pdf, accessed October 20, 2020.

MISE (2020) "Proposte per una Strategia italiana per l'intelligenza artificiale". Italian Ministry of Economic Development. Rome, Ministero per lo Sviluppo Economico.

Osservatori.net (2019). "Il mercato e le applicazioni di artificial intelligence: lo scenario italiano", AI Observatory of the Polytechnic of Milan: Osservatorio Intelligenza Artificiale. Milan: Politecnico School of Management. Available at <https://www.osservatori.net/it/prodotti/formato/video/mercato-applicazioni-artificial-intelligence-italia-2018>,

The Italian Prime Minister and the Minister for Technological Innovation and Digitalisation (2020). "Decreto del Presidente del Consiglio dei Ministri". Rome, July. Available at https://pianotriennale-ict.italia.it/assets/pdf/2020-2022/DPCM_17_luglio_2020_pdf_testo.pdf, accessed October 20, 2020.

The President of the Republic of Italy (2020). "Urgent measures for simplification and digital innovation". Decreto-Legge, 16 luglio 2020, n. 76. Official Gazette, Rome. Available at <https://www.gazzettaufficiale.it/eli/gu/2020/09/14/228/so/33/sg/pdf>, accessed October 21, 2020.

AI policy in the Czech Republic

Strong business focus, welcoming
towards foreign investment

Anas Zaza

Anas Zaza is an External Analyst with the Institute of Politics and Society in the Czech Republic and completing studies in International Relations at Anglo-American University in Prague. He is also an Analyst and Contract Specialist with Runecast Solutions and an independent researcher.

Summary

AS AI SYSTEMS continue to impact society and citizens, national AI strategies are attempting to capture economic prosperity and remain competitive in the 21st century. In the world of AI, keywords such as algorithms, automation and big data are increasingly becoming commonplace. They are primarily driven by an AI resurgence in the private sector that has seeped into all aspects of society.

This chapter analyses the Czech Republic's National Artificial Intelligence Strategy (NAIS), published in May 2019, and how or to what extent the strategy is influenced by the European Union's approach on "trustworthy AI". This notion is emanating from the appointment of the High-Level Expert Group on AI (AI HLEG) and its key publications,

primarily the *Ethics Guidelines for Trustworthy AI*. Specifically, the seven key requirements for the realisation of trustworthy AI, found in the Ethics guidelines, as well as other documents including similar notions, from the OECD AI Policy Observatory¹ and European Commission's AI Watch², are here part of the analytical foundation.

As concluded in this chapter, the NAIS – being published in the month after the Ethics guidelines – has not explicitly linked the seven key requirements, but can be seen in light of the preceding European AI strategy (April 2018) that is echoed in the Communication from April 2019. The Communication puts forward a human-centric approach to AI, whereby “AI is not an end in itself, but a tool that has to serve people with the ultimate aim of increasing human well-being”.³ The NAIS encompasses nearly all aspects indirectly with the aim of improving the country's economic growth and competitiveness in AI by creating favourable policy conditions.

The NAIS provides a list of tools and tasks and the main coordinating role is assigned to the Ministry of Industry and Trade with the aim that the strategy is to be continually reviewed, on a yearly basis, with reports on the fulfilment of objectives and proposals to revise objectives and instruments.

As the strategy takes shape, an AI Committee has been established to supervise its implementation. The strategy is compatible with previous industry policy documents such as Industry 4.0 (Průmysl 4.0), a national initiative that aimed at maintaining and enhancing the competitiveness of the Czech Republic in the wake of the so-called Fourth Industrial Revolution.⁴ More importantly, the NAIS is explicitly referring to the official EU strategies, while significant salience is also given to the national *Innovation Strategy of the Czech Republic 2019–2030*, as the

1 OECD (2020).

2 AI Watch (2020).

3 European Commission (2019).

4 Digital Transformation Monitor (2017).

stated intent is to complement and expand upon it, alongside the similarly important government strategy *Digital Economy and Society*.

1. Analysis

ADVANCES IN THE application of artificial intelligence (AI) systems have gained momentum in the past decade, bringing forth national initiatives and strategies to capture competitive advantages of economic growth and beneficial societal impacts. Member States were “encouraged to develop their national AI strategy by mid-2019, building on the work done at the European level”, as called for in the Coordinated Plan on Artificial Intelligence from December 2018.

The European Strategy on Artificial Intelligence is led by the AI HLEG – composed of appointed experts by the European Commission – to present recommendations on future-related policy developments and on ethical, legal and societal issues related to AI, including socio-economic challenges.⁵ The Ethics Guidelines for Trustworthy AI were published by the AI HLEG in April 2019, along with 33 proposed Policy and Investment recommendations for Trustworthy AI. These documents were addressed to EU institutions and Member States and were published in June 2019; additional emphasis on building an “ecosystem of trust” appeared in the Commission’s White Paper on Artificial Intelligence published in February 2020.

Although the Czech Republic’s NAIS was released in May 2019,⁶ only a month after *The Ethics Guidelines*, the NAIS draws reference to guidelines and publications that align with the pan-European approach and is inspired by similar foreign strategic documents.

The NAIS was created in collaboration with a diverse team of experts from academia, as well as the private sector. Among them are representatives of ministries, institutions and experts from the Academy of

⁵ AI HLEG (2019).

⁶ European Commission (2020).

Sciences of the Czech Republic.⁷ The stated goal is for the government, public and private sectors to coordinate closely as AI systems improve, while meeting the objectives of the Government Innovative Strategy 2019-2030 and the Digital Czech Republic programme.⁸

The AI Committee that is assigned through the NAIS is intended to be a subcommittee of the Steering Committee of the Digital Czech Republic strategy, chaired by the Deputy Minister of Industry and Trade for Digitisation and Innovation, with interim reports to be submitted yearly to the Steering Committee and to the government, to inform about progress of strategy implementation. For some of the seven different strands of AI policy that the report outlines (see below), other government departments are recruited to act as coordinating members: the Ministry of Education, Youth and Sports; the Ministry of Labour and Social Affairs; the Office of the Government of the Czech Republic. It is through this newly established AI Committee that the responsible ministers (the Deputy Prime Minister and the Minister of Industry and Trade) are implementing and coordinating the strategy as a whole. The operational management of the AI Committee will be continuously ensured by its Executive Committee.⁹

1.1. Composition of the NAIS

The NAIS is divided into seven chapters, and its key areas are Research and Development (R&D), financing, industry, human capital, education, regulation, and international cooperation. For each chapter, the NAIS stipulates a responsible Ministry as Coordinator, a baseline providing the current state, the policy initiatives that will be developed, the cooperating entities, tools and methodology, and the key objectives that are targeted at short-term (until 2021), medium-term (until 2027) and long-term (until 2035).¹⁰

⁷ NAIS (2019).

⁸ Government of the Czech Republic (2019).

⁹ European Commission (2020).

¹⁰ European Commission (2020).

The NAIS also presents results from an empirical mapping of the industrial landscape, intended to ground the report in actual data and knowledge from practice. Hence, the annex includes results from a survey that was conducted among organisations and companies engaged in AI activities, in both academia and the private sector. This survey was led by the Confederation of Industry of the Czech Republic, who are responsible for making the initial evaluations of the corporate AI environment. “The members of the Confederation of Industry of the Czech Republic, all other member associations and organisations and the general public were approached for the purpose of mapping – in total thousands of entities were approached, of which 50 companies decided to become involved at this stage.”¹¹ This quantitative mapping gives indicative data on the applications and development areas of AI, methods used in AI, the sectors of integration of AI, and the distribution across various industries in the Czech Republic. The “application development areas of AI” that were most often mentioned were “information and communication activities” and “business process support,” and the sectors that were most often mentioned were corporate AI and software development. 44 percent of participants highlighted the importance of education and retraining of workers, which the NAIS also covers in its fourth chapter, which addresses human capital and the education system and lifelong learning, and also its fifth chapter, which addresses the impacts of AI on the labour market and social systems.

1.2. Focus areas of the NAIS

One of the primary focuses stated in the report is supporting R&D in AI, where the transnational ambition to build European Centres of Excellence in AI research, Test Centres, and Digital Innovation Hubs are given clear prominence in the report.¹² The prospective establishment of a European Centre of Excellence in AI, based on a consortium of academic research institutes, is outlined to be based in the capital (Prague) and to reach throughout the Czech Republic. This is concordant with the AI HLEG’s recommendations on Policy and Investment Recommendations which highlights the aim of ensuring world class

¹¹ NAIS (2019), p. 43.

¹² NAIS (2019).

research capabilities in AI within the EU. The NAIS plans integrated systems of transfer of academic know-how into the EU well into 2035, with the goal of making the Czech Republic an attractive AI research destination that would be connected with other centres throughout the EU further expanding academic research on an international stage. This is fulfilled by deepening cooperation with global AI centres, as well as maintaining top research and experts in the Czech Republic.

The keywords of the report seem to be competitiveness, R&D, digital infrastructure, skills and training, and commercially viable innovation. The main objective appears to be the concentration of excellent R&D in AI, in particular by supporting the creation of the European Centre of Excellence, Test Centre and Digital Innovation Hubs. The report thus outlines seven strands that are co-constitutive of such an agenda:

- ▶ Conditions for attracting top foreign talent (1)
- ▶ Funding for research, the development of start-ups, SMEs (2)
- ▶ Provision of digital infrastructure (3)
- ▶ Education and lifelong learning; both technical and humanities-oriented (4)
- ▶ Adaptable social welfare system (5)
- ▶ Legislation; ensuring the protection of fundamental rights and security as well as legal certainty for investors (6)
- ▶ International cooperation and involvement, especially at the EU level (7)

For each of these seven strands, the report outlines a set of tools, partners, and short-, medium-, and long-term objectives.

In the listed strategic documents, the publications and guidelines of the AI HLEG are not present, but reference to communications from the European Commission on “Coordinated Plan on Artificial Intelligence in Europe”,¹³ “Artificial Intelligence for Europe”,¹⁴ and “A Digital Single Market Strategy for Europe”.¹⁵

13 European Commission (2018a).

14 European Commission (2018b).

15 European Commission (2015).

The NAIS is well-structured and clear objectives are laid out through the seven vertically divided chapters. The overall aim of the strategy is to improve the country's economic growth and competitiveness in AI by creating favourable policy conditions, “all while maintaining a high level of protection of fundamental and other rights and in the line with the European approach of human-centric AI.”¹⁶

1.3. Law and ethics

The Ethics Guidelines by the AI HLEG states that AI systems need to be human-centric on the basis of a “commitment to their use in the service of humanity and the common good” towards a Trustworthy AI.¹⁷ The guidelines in creating Trustworthy AI include three components of lawful, ethical and robust systems all intended to ensure a Trustworthy AI framework.

Among many of the NAIS's short-term objectives, up until 2021, is the establishment of an Expert Platform and Forum (modelled on the “Observatory and Forum” set up by the EU Commission) in cooperation with the Institute of State and Law of the Academy of Sciences of the Czech Republic, for the continuous monitoring of legal and ethical rules at both a national and international level, adhering to the human centric AI.

The NAIS has taken measures in assuring national and internal laws as well as ethical guidelines, when it comes to either creating legislation, ensuring lawful development and implementation of AI systems, or by removing legal barriers to AI development, including public law, also cited public procurement.

Consequently, the report's sixth strand, *Legal and societal aspects of AI, ethical rules, consumer protection and security issues*, provides a pathway and a commitment to following European Union Law. What is requested is clear legislation, ensuring the protection of fundamental rights and security as well as legal certainty for investors and citizens

¹⁶ NAIS (2019).

¹⁷ AI HLEG (2019).

alike, as well as means to safeguard a democratic development. What is sought is therefore “the creation of an administrative and legislative framework for AI that avoids any form of discrimination or disadvantage, with a strong emphasis on rights and privacy”.¹⁸ In the stated measures to address the impacts of the AI on the labour market and the social system (the fifth strand, seen to be coordinated by the Ministry of Labour and Social Affairs), potential threats are identified, include:

[Deepening] problems in socially excluded regions, temporarily increasing structural and frictional unemployment, or [...] various forms of inequality and discrimination. The effects of automation are likely to be different for different population groups, with the middle class being among the most affected groups. The impacts can thus be not only purely economic, but also social and political.¹⁹

1.4. AI, data and digitisation

Although unclear as to the adherence to the Trustworthy AI guidelines, the NAIS implements a European approach, and aims to fulfil the strategy of Digital Czech Republic. At a national level, this includes a myriad of major government strategies for digitalisation and innovation, most comprehensively outlined in the Innovation Strategy of the Czech Republic 2019-2030. The Digital Czech Republic international conference is held annually, since 2015, and brings together speakers across diverse panels geared towards discussion and debate around policy programmes such as AI Citizens Safety, the future of jobs, the future of education, data rights and digitalisation amongst a few other areas and topics.

The four ethical principles forming a foundation for Trustworthy AI, such as the respect for human autonomy, prevention of harm, fairness and explicability are taken into consideration in the implementation of the NAIS. However, these four objectives are not stated explicitly, meaning that there is more room for further expansion of the AI HLEG

¹⁸ NAIS (2019), p.8.

¹⁹ Ibid., p.30.

guidelines. Areas such as Privacy and Data Governance (addressing the opportunities for individuals to trust the data processing, ensuring adequate civic control over data, and that data not be used to harm or discriminate against) could be further expanded in future iterations of the AI policy of the Czech Republic.

1.5. Timeline

The NAIS lists both medium-term (timeframe 2027) and long-term (timeframe 2035) objectives. The medium term objectives propose plans on implementing AI development and usage tools in accordance with ethical and legal rules (including Ethical Guidelines for Artificial Intelligence Development and Use) and human centric AI.²⁰ Additionally, the sixth strand of the NAIS explicitly states that the most important influence is that of the European Union legal framework, through which strategic goals in the field of AI (including the creation of Ethical Guidelines for the Development and Use of AI) are already under way. Also in this legal framing of AI, the NAIS makes particular reference to »innovation-friendly« aspects of regulation. Other international organisations, notably the OECD, the WTO, the UN and the Council of Europe, are noted to have significant activity in this field as well.

While the NAIS appears to lack concrete arguments regarding things like civic accountability of AI Systems, as per *The Ethics Guidelines*, the strategy stipulates the need for “clear and timely AI regulation to ensure legal certainty for citizens, entrepreneurs and investors”.²¹ Moreover, the NAIS touches upon the support of certified methodology of implementation and specialised courses, including the proposal of a certified methodology for system audits in co-operation of the public and private sectors.²²

²⁰ NAIS (2019).

²¹ NAIS (2019), p.36.

²² NAIS (2019).

2. Conclusion

THE NAIS IS part of the implementation of the Innovation Strategy of the Czech Republic 2019-2030 and lays out the foundation for building a comprehensive ecosystem intended to answer to the stated aims within the EU to create European Centres of Excellence. In this framework, the design, creation, implementation and successful promotion – coupled with favourably investing opportunities and innovative management – of Digital Innovation Hubs (DIH) provide one of the links between the public and digitalisation. Highlighted in the executive summary of the NAIS:

It is the cooperation of all the entities involved that is crucial for the real fulfillment of the National AI Strategy and successful handling of fundamental changes for the Czech economy and society. Ensuring equal opportunities and the economic development for the entire society within Europe and internationally.²³

However, when analysed in line with the EU Commission approach on “trustworthy AI”, the NAIS has not explicitly linked the seven key requirements with the AI HLEG delivered Ethics Guidelines on Artificial Intelligence, putting forward a human-centric approach to AI, whereby “AI is not an end in itself, but a tool that has to serve people with the ultimate aim of increasing human well-being”.²⁴

3. Recommendations

WITH THE ANTICIPATED yearly reports on the fulfilment of the strategy of the NAIS objectives, which is to be submitted to the Steering Committee of the Digital Czech Republic Strategy and the Government of the Czech Republic, proposals are likely to be presented by the committee so as to revise objectives and instruments continually.

²³ Ibid.

²⁴ European Commission (2019).

A worthy expansion of such an endeavour would be to emphasise *human agency and oversight*, which helps ensure that AI systems do not undermine human autonomy or cause other adverse effects. Depending on the specific AI-based system and its application area, the appropriate degrees of control measures, including the adaptability, accuracy and explainability of AI-based systems, should be explained further in a series of supporting documents.

The NAIS could further elaborate also on the *technical robustness and safety* of AI systems, as regards their intended reliability. AI systems ought to be secure enough to be resilient against both overt attacks and more subtle attempts to manipulate data or algorithms themselves, and ensure fall-back plans in case of problems. The decisions delivered by such systems are to be accurate, or at least correctly reflect their level of accuracy, and their outcomes should be reproducible.

The *traceability* of AI systems should be ensured, as it is important to log and document both the decisions made by the systems, as well as the entire process that yielded the decisions. In addition, future iterations of the NAIS could provide more thorough explanations of the degree to which AI systems are thought to influence and shape organisational decision-making processes.

In a ranking presented by Oxford Insights and the International Research Development Centre (IDRC), drawing upon 33 indicators, the Czech Republic scored 61.5 (out of 100) in the Government AI Readiness Index 2020, ranking 32/172 internationally and 4/20 in regional rank.²⁵ Ambitiously, the Czech Republic aims to be “the model country for the whole of Europe in automation before the Czech Republic’s presidency of the EU Council in 2022,” as highlighted in prime minister Andrej Babiš’s “Vision – Czech Republic, the country of robots,” in the opening pages of the strategy,²⁶ and the intent seems to be that the objectives and timelines set forth in the NAIS will be accommodating to this.

²⁵ Oxford Insights (2020).

²⁶ NAIS (2019), p.4.

Importantly, the value from the linking of the NAIS, in its current form (released in May 2019), to the AI HLEG and other publications past this date, would help accommodating objectives for future review and amendment of the Czech strategy, helping to ensure that the Czech Republic follows guidelines set forth at the European level as highlighted in the sixth chapter of the NAIS.

References

AI HLEG (2019). *Ethics Guidelines for Trustworthy AI*. Brussels. European Commission: High-Level Expert Group on AI. April 8. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

AI Watch (2020). *European Commission: Knowledge for policy*. https://ec.europa.eu/knowledge4policy/ai-watch_en

Digital Transformation Monitor (2017). Czech Republic: “Průmysl 4.0”. European Commission, May. [https://ec.europa.eu/growth/tools-databases/dem/monitor/content/czech-republic-%E2%80%99C5%AFmysl-40%E2%80%9DEuropean Commission \(2015\). “Communication: A Digital Single Market Strategy for Europe” \(COM\(2015\) 192 final\). May 5. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52015DC0192](https://ec.europa.eu/growth/tools-databases/dem/monitor/content/czech-republic-%E2%80%99C5%AFmysl-40%E2%80%9DEuropean Commission (2015). “Communication: A Digital Single Market Strategy for Europe” (COM(2015) 192 final). May 5. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52015DC0192)

European Commission (2018a). “Communication: Artificial Intelligence for Europe” (COM(2018) 237 final). April 25. <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>

European Commission (2019). “Communication: Building Trust in Human Centric Artificial Intelligence” (COM(2019)168). April 8. <https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence>

European Commission (2018b). “Communication: Coordinated Plan on Artificial Intelligence” (COM(2018) 795 final). European Commission. December 7. https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en

European Commission (2020). *Czech Republic AI Strategy Report - Knowledge for policy* European Commission. Last updated: August 5. European Commission: Knowledge for Policy. https://ec.europa.eu/knowledge4policy/ai-watch/czech-republic-ai-strategy-report_en

Government of the Czech Republic (2019). *Innovation Strategy of the Czech Republic 2019-2030. The Country For The Future*. <https://www.vyzkum.cz/FrontAktualita.aspx?aktualita=867990>

NAIS (2019). *National Artificial Intelligence Strategy of the Czech Republic*. Prague: Ministry of Industry and Trade of the Czech Republic. https://www.mpo.cz/assets/en/guidepost/for-the-media/press-releases/2019/5/NAIS_eng_web.pdf

OECD (2020). *OECD AI Policy Observatory*. <https://www.oecd.ai/>

Oxford Insights (2020). *Government AI Readiness Index 2020*. Commissioned by Canada’s International Development Research Centre (IDRC). <https://www.oxfordinsights.com/government-ai-readiness-index-2020>

Human-Centred AI in the EU

The European approach to artificial intelligence (AI) points to ethical considerations, human control and trustworthiness as its core tenets. But how clearly is this approach reflected in the Member States' strategies?

This anthology analyses to what extent the notions of ethical and trustworthy AI, presented by the High-Level Expert Group on Artificial Intelligence and the European Commission, have influenced AI strategies in Portugal, The Netherlands, Italy, the Czech Republic, Poland, Norway as well as the Nordics overall.

It is clear that the EU-level policies have had an impact on the national level strategies, although sometimes only to the extent that they were published before the national documents. For instance, while some countries, such as Norway and Portugal, have explicitly incorporated aspects from the Ethics Guidelines, others, such as the Nordics, already tended to include questions of trust and transparency, or on ethics as in the case of Poland.

The EU has emphasised AI trustworthiness as both an ethical imperative and competitive advantage. However, implementation is still at the starting line: much depends on alignment between this diverse group of nations, with different priorities, within the single market.

Editors

Stefan Larsson
Claire Ingram Bogusz
Jonas Andersson Schwarz

Commentary

Fredrik Heintz

Authors

Pedro Rubim Borges Fortes
Francesco Cappelletti
Frans af Malmborg
Stephen Cory Robinson
Kasia Söderlund
Katja de Vries
Anas Zaza