verbraucherzentrale
Bundesverband

UNIVERSITY OF AMSTERDAM
Institute for Information Law

Kristina Irion

AI REGULATION IN THE EUROPEAN UNION AND TRADE LAW:

HOW CAN ACCOUNTABILITY OF AI AND A HIGH LEVEL OF CONSUMER PROTECTION PREVAIL OVER A TRADE DISCIPLINE ON SOURCE CODE?

26 January 2021

Gefördert durch:

Bundesministerium der Justiz und für Verbraucherschutz

aufgrund eines Beschlusses des Deutschen Bundestages

Impressum

Verbraucherzentrale Bundesverband e.V.

Team Brussels Office

Rudi-Dutschke-Straße 17

10969 Berlin

buero-bruessel@vzbv.de

Bundesverband der Verbraucherzentralen und Verbraucherverbände

Verbraucherzentrale Bundesverband e.V.

SUMMARY

Artificial Intelligence (AI) applications can bring many benefits for consumers, as well as influence consumer behaviour and the choices they make. On a large scale AI can profoundly transform consumer markets by, for example, enabling fully personalised consumer transactions on a population-wide scale. AI-powered consumer services rapidly diffuse across the global digital ecosystem thereby connecting consumers in the European Union (EU) to business operating from outside the EU. Individuals who are at the receiving end of AI systems must be reassured that these technologies operate in a way that respects fundamental and consumer rights.

The Federation of German Consumer Organisations (Verbraucherzentrale Bundesverband – vzbv) has commissioned this study from the Institute for Information Law (IViR) at the University of Amsterdam, in order to shed light on the cross-border supply of AI technology and its impact on EU consumer rights.[1]

In the current negotiations on electronic commerce at the World Trade Organisation (WTO), the EU supports the introduction – in the legal text – of a clause which prohibits the participating countries to introduce – in their national laws – measures that require access to, or transfer of, the source code of software, with some exceptions. This is a cause for concern for experts and rights advocates, as such a clause – if not carefully conditioned – can prevent future EU regulation of AI that may be harmful to consumers.

This study concludes that the source code clause within trade law indeed restricts the EU's right to regulate in the field of AI governance in several important ways.

The conclusion is surprising given that EU trade policy documents make no reference to AI, only to electronic commerce, and that no direct link has been made between the clause on software source code and algorithms. This study raises an important EU policy issue that deserves to be put to democratic scrutiny and discussion before the EU agrees to a new clause on software source code in a plurilateral WTO agreement on electronic commerce.

This study forms a comprehensive understanding of this issue that intersects three different areas: (1) emerging EU governance of AI and (2) the application of EU consumer protection law to AI with (3) the EU's position in the WTO electronic commerce negotiations.

The General Agreement on Trade in Services (GATS), to which the EU is a member, already applies to cross-border trade in AI-powered digital services. What is more, computer and machine learning algorithms are expressed in source code and would thus be protected under the new trade law clause on software source code. A legislation requiring auditing of source code ("white box" method) but also auditing of inputs and outputs of an AI system via its interfaces ("black box" method) violates the trade law clause on source code. A trade law violating legislation can be justified pursuant to the GATS exceptions if the requirements that are attached to the exceptions can be satisfied.

This means that the EU's possibility to adopt rules that, for example, mandate external audits of AI systems will be confined to the policy space that is allowed under trade law.

According to Article 207(3) TFEU, the Council and the European Commission are responsible for ensuring that trade agreements are compatible with internal Union policies

[1] This research has been conducted in full compliance with the 2017 European Code of Conduct for Research Integrity. See ALLEA - All European Academies, 'The European Code of Conduct for Research Integrity' (2017) <https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf>.

and rules. Several policy options for AI governance that are currently discussed at the EU level risk being inconsistent with the WTO electronic commerce proposal on source code, unless they neatly fit the GATS general exceptions. For example:

- The **European Commission's White Paper on AI** proposes the introduction of prior conformity assessment of high-risk AI applications by certified testing centres;[2]
- **Germany's Data Ethics Commission** recommends "always-on" regulatory oversight of algorithmic systems with a high potential for harm through a live interface;[3]
- The **Digital Services Act** proposal requires very large online platforms to enable vetted researchers to study systemic risks by accessing data via interfaces (APIs).[4]

Already today technical interfaces (public-facing or internal APIs) are of strategic importance for ensuring accountable and trustworthy AI.[5] Committing to a trade law clause that would make it harder to engage with AI systems via these interfaces or mandate standardized interfaces in the interest of auditability is counterproductive.

Another area for conflict between EU policy and trade law arises where a high level of consumer protection calls for robust safeguards against anticipated risks of AI technology. A recurring theme is AI systems' characteristic opacity and the difficulty of proving that an AI system is faulty, biased or unfair. Defending consumer rights in digital consumer markets requires more agile and scalable regulatory measures, in addition to the current system of ex post enforcement.

Monitoring the effects of AI systems in digital consumer markets would benefit from regulation enabling effective external audits of AI systems. Such regulation would mandate accountability via external audits of the input data and outputs from an AI system and the setting up of auditing interfaces in order to verify that EU consumer rights are complied with. The source code clause in trade agreements, by contrast, would not only protect computer and machine learning algorithms but also the interfaces of an AI system that are indispensable for audits.

It is important to note also that digitalization leads to more and more digital artefacts made of software source code, and AI technology may give rise to new risks for individuals and society whilst trade law largely remains static after having been ratified. The source code clause is too broad for domestic digital policies that need to build on interoperability, accountability, and verifiability of digital technologies.

IN LIGHT OF THIS, THE STUDY RECOMMENDS TWO OPTIONS:

1. The European Commission should clarify the impact of the source code clause on EU digital policies, in particular consumer rights, and meanwhile give up on this trade law clause since software source code already enjoys copyright and trade secret protection; or
2. The European Commission should limit the trade law clause to the situation of forced technology transfers for dishonest commercial practices, or carve out measures on algorithmic accountability from the scope. This would be prudent and provide time to develop robust domestic policy as well as international standards on accountable AI.

---

[2] See European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)' 23 <https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf>.

[3] See Data Ethics Commission, 'Opinion of the Data Ethics Commission' (2019) 184 <https://datenethikkommission.de/wp-content/uploads/DEK_Gutachten_engl_bf_200121.pdf>.

[4] See European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM/2020/825 final) <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608117147218&uri=COM%3A2020%3A825%3AFIN>

[5] See Section II.2.3 on Interface audit and Section III.1.4 on the Opinion of the Data Ethics Commission.

ZUSAMMENFASSUNG

Anwendungen der Künstlichen Intelligenz (KI) können Verbrauchern viele Vorteile bieten, gleichzeitig können sie aber auch das Verbraucherverhalten und Entscheidungen maßgeblich beeinflussen. Im großen Maßstab kann KI Verbrauchermärkte grundlegend verändern, indem etwa personalisierte Angebote gesamtgesellschaftlich ermöglicht werden. KI-gestützte Dienstleistungen verbreiten sich schnell im globalen digitalen Ökosystem und verbinden europäische Verbraucher mit Unternehmen in der ganzen Welt. Verbraucher, die durch KI-Systeme beurteilt werden, müssen darum sicher sein, dass diese Technologien die Grund- und Verbraucherrechte respektieren.

Die Verbraucherzentrale Bundesverband (vzbv) hat diese Studie beim Institut für Informationsrecht (IViR) der Universität Amsterdam in Auftrag gegeben, um die grenzüberschreitende Nutzung von KI-Technologien und deren Auswirkungen auf die Verbraucherrechte in der EU zu beleuchten.

In den aktuellen Verhandlungen über den elektronischen Handel unter dem Dach der Welthandelsorganisation (WTO) unterstützt die EU die Aufnahme einer Klausel, die es den teilnehmenden Ländern verbietet, in ihren nationalen Gesetzen Maßnahmen einzuführen, die den Zugang zu oder die Weitergabe von Software-Quellcodes vorschreiben – mit einigen Ausnahmen. Dies gibt in der Zivilgesellschaft Anlass zur Sorge, dass eine solche Klausel – wenn sie nicht sorgfältig abgegrenzt ist – eine zukünftige Regulierung von KI erschweren kann, was für Verbraucher schädlich sein könnte.

Diese Studie kommt zu dem Schluss, dass die Quellcode-Klausel im Handelsrecht das Recht der EU, im Bereich der KI-Regulierung zu regulieren, in mehreren wichtigen Punkten einschränkt.

Die Schlussfolgerung kommt überraschend, da in den handelspolitischen Dokumenten der EU kein Bezug zu KI, sondern nur zum elektronischen Handel genommen wird und kein direkter Zusammenhang zwischen der Klausel über Software-Quellcode und Algorithmen und ihrer Regulierung hergestellt wird. Die Studie betrachtet ein wichtiges europapolitisches Thema, das einer demokratischen Prüfung und öffentlicher Diskussion unterzogen werden sollte, bevor die EU einer neuen Klausel über Software-Quellcode in einem plurilateralen WTO-Abkommen über elektronischen Handel zustimmt.

Dieser Studie liegt ein umfassendes Verständnis dieses Themas zu Grunde, das drei Bereiche betrachtet: (1) die sich abzeichnende EU-Regulierung von KI und (2) die Anwendung des EU-Verbraucherschutzrechts auf KI-Anwendungen mit (3) der Position der EU in den WTO-Verhandlungen zum elektronischen Handel.

Das Allgemeine Abkommen über den Handel mit Dienstleistungen (GATS), dem die EU angehört, gilt bereits für den grenzüberschreitenden Handel mit KI-gestützten digitalen Dienstleistungen. Darüber hinaus werden Computeralgorithmen, auch solche im Bereich KI, in Quellcode ausgedrückt und wären somit durch die neue Handelsrechtsklausel für Software-Quellcode geschützt. Eine Gesetzgebung, die nicht nur die Prüfung des Quellcodes ("White-Box"-Methode), sondern auch die Input-Output-Analyse eines KI-Systems über seine Schnittstellen ("Black-Box"-Methode) vorschreibt, verstößt gegen die handelsrechtliche Klausel zum Quellcode. Eine gegen das Handelsrecht verstoßende Gesetzgebung kann gemäß den GATS-Ausnahmen gerechtfertigt werden, sofern die rechtlichen Voraussetzungen, an die diese Ausnahmen geknüpft sind, erfüllt werden können.

Das bedeutet, dass die Möglichkeit der EU Vorschriften zu erlassen, die externe Audits von KI-Systemen vorschreiben, auf den handelsrechtlich zulässigen Spielraum beschränkt sein würden.

Gemäß Artikel 207(3) AEU-Vertrag sind der Rat und die Europäische Kommission dafür verantwortlich, dass Handelsabkommen mit der internen Politik und den internen Vorschriften der Union vereinbar sind. Mehrere Regulierungsoptionen für die KI-Regulierung, die derzeit auf EU-Ebene diskutiert werden, laufen Gefahr, mit dem WTO-Vorschlag der EU zum elektronischen Handel unvereinbar zu sein. Es sei denn, sie können gemäß den allgemeinen GATS-Ausnahmen gerechtfertigt werden. Zum Beispiel:

- Das **Weißbuch der Europäischen Kommission zu KI** schlägt die Einführung einer vorgeschalteten Konformitätsbewertung von KI-Anwendungen mit hohem Risikopotenzial durch zertifizierte Prüfstellen vor;
- Die deutsche **Datenethikkommission** empfiehlt eine Live-Schnittstelle zur kontinuierlichen Kontrolle von algorithmischen Systeme mit erheblichem Schädigungspotenzial durch eine Aufsichtsbehörde; oder
- Der Vorschlag für einen **Digital Services Act** der Europäischen Kommission verlangt in Bezug auf sehr große Online-Plattformen, dass es zugelassenen Wissenschaftlern ermöglicht wird, über Schnittstellen (APIs) auf Daten zuzugreifen, um systemische Risiken zu untersuchen.

Bereits heute sind technische Schnittstellen (öffentlich zugängliche, oder interne APIs) von strategischer Bedeutung, um verantwortliche und vertrauenswürdige KI zu gewährleisten. Sich auf eine Handelsrechtsklausel zu verpflichten, die den Umgang mit KI-Systemen über diese Schnittstellen oder die Einführung standardisierter Schnittstellen zur Stärkung einer Überprüfbarkeit erschwert, ist kontraproduktiv.

Ein weiteres Konfliktfeld zwischen EU-Politik und Handelsrecht ergibt sich dort, wo ein hohes Maß an Verbraucherschutz robuste Schutzmaßnahmen gegen die zu erwartenden Risiken von KI-Technologien erfordert. Das gilt besonders für die charakteristische Intransparenz von KI-Systemen und die Schwierigkeit zu beweisen, dass ein KI-System fehlerhaft, voreingenommen oder unfair ist. Der Schutz von Verbraucherrechten in digitalen Verbrauchermärkten erfordert skalierbare Regulierungsmaßnahmen, zusätzlich zum bestehenden System der nachträglichen Rechtsdurchsetzung.

Die Überprüfbarkeit von KI-Systemen mit dem Ziel des Verbraucherschutzes in der digitalen Welt, kann insbesondere durch externe Audits von KI-Systemen ermöglicht werden. Eine entsprechende Regulierung würde eine Rechenschaftspflicht durch externe Audits von Eingabedaten und Ausgaben eines KI-Systems und der Einrichtung von Audit-Schnittstellen profitieren. So könnte überprüft werden, ob europäisches Verbraucherrecht eingehalten wird. Die Quellcode-Klausel in Handelsabkommen würde dagegen nicht nur Computer- und Machine-Learning-Algorithmen schützen, sondern auch die Schnittstellen eines KI-Systems, die für solche externe Überprüfungen unerlässlich sind.

Zudem ist zu beachten, dass die Digitalisierung zu immer mehr digitalen Artefakten aus Software-Quellcode führt und die KI-Technologie neue Risiken für Individuen und die Gesellschaft mit sich bringen kann. Das Handelsrecht gleichzeitig aber nach seiner Ratifizierung weitgehend statisch bleibt und die Hürden für eine Anpassung hoch sind. Die Quellcode-Klausel ist zu weit gefasst für eine nationale und europäische Digitalpolitik, die auf Interoperabilität, Verantwortlichkeit und Überprüfbarkeit digitaler Technologien aufbauen muss.

VOR DIESEM HINTERGRUND EMPFIEHLT DIE STUDIE ZWEI OPTIONEN:

1. Die Europäische Kommission sollte die Auswirkungen der Quellcode-Klausel auf die Digitalpolitik der EU, insbesondere auf den Verbraucherschutz, klarstellen. In der Zwischenzeit sollte auf diese handelsrechtliche Klausel verzichtet werden, da Software-Quellcode nach wie vor im Handelsrecht Urheberrechtsschutz und den Schutz von Geschäftsgeheimnissen genießt; oder
2. Die Europäische Kommission sollte die Handelsrechtsklausel auf Fragen des erzwungenen Technologietransfers für unlautere Geschäftspraktiken beschränken, oder Maßnahmen zur Algorithmenkontrolle deutlich aus dem Anwendungsbereich ausklammern. Dies wäre umsichtig und würde Zeit verschaffen, um eine robuste nationale und europäische Politik und internationale Standards für verantwortliche KI zu entwickeln.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADM	Algorithmic decision making
AI	Artificial intelligence
API	Application Programming Interface
CPTPP	Comprehensive and Progressive Agreement for Trans-Pacific Partnership
EU	European Union
GATS	(WTO) General Agreement on Trade in Services
GDPR	(EU) General Data Protection Regulation
ML	Machine learning
OECD	Organization for Economic Cooperation and Development
RCEP	Regional Comprehensive Economic Partnership
TFEU	(EU) Treaty on the Functioning of the European Union
USMCA	United States-Mexico-Canada Agreement
VPA	Virtual Personal Assistant
WTO	World Trade Organisation

INTRODUCTION

Organisations in the public and private sector enthusiastically embrace digital technologies that can automate routine tasks, manage complex workflows, and recognise patterns in copious amounts of digital data on the basis of which predictions and decisions can be made. The state-of-the-art technology that makes this possible is called Artificial Intelligence (AI) which enables computers to learn from data. AI technology is routinely being used today in digital applications and services that underpin the relationships and transactions between individuals and government, or individuals and businesses.

Emerging AI applications can bring many benefits for consumers, influence consumers' behaviour and the way they make choices. On a large scale AI can profoundly transform consumer markets by for example enabling fully personalized consumer transactions on a population-wide scale. Individuals and consumers who are at the receiving end of AI systems must be reassured that these technologies are implemented and operate in compliance with EU fundamental rights and the body of consumer protection laws. New challenges arise from AI technologies' opacity and scalability which for example facilitates unprecedented mass-personalisation in consumer markets.

An additional layer of complexity stems from the fact that AI can be applied across borders. In today's digital ecosystem, it is quite common that EU consumers use digital services that incorporate AI technology in their software architecture and are supplied by businesses outside the EU. To shed light on the cross-border supply of AI technology and its bearing on EU consumer rights the Federation of German Consumer Organisations (Vzbv) has commissioned this study from the Institute for Information Law (IViR) at the University of Amsterdam. This research has been conducted in full compliance with the 2017 European Code of Conduct for Research Integrity.[6]

The objective of this study is to explore how a high level of consumer protection can be attained in the context of AI applications supplied to consumers from outside the EU. The study compares the EU approach to the governance of AI in light of EU consumer rights with the EU proposal in the World Trade Organization's (WTO) negotiations on electronic commerce.[7] The EU proposal contains a new discipline that prohibit member states to require the transfer of, or access to, source code of software. A particular focus of this study will be on the implications of such a source code discipline for current and future EU policy that aims to ensure transparent and accountable AI.

This study aims to generate the understanding required to judge whether a new source code discipline inside trade law would curtail the EU approach to algorithmic transparency and accountability, and consequently be detrimental to consumer rights. To this end, the **study assesses the internal consistency of EU policies across three legal domains** with a focus on safeguarding consumer rights in cross-border commerce involving AI:

1. Emerging EU governance of AI,
2. AI risks anticipated for consumer rights, and
3. the EU position in the WTO electronic commerce negotiations.

All three legal domains are currently evolving, the first two exclusively at the level of the EU, whereas the third concerns an EU position in the plurilateral negotiations at the WTO. The governance of AI systems is still on the drawing board with EU policymakers who are preparing a legislative proposal planned for early 2021. It is not yet clear how the existing body of EU consumer rights will interface with a new AI regulation, for instance

[6] ALLEA - All European Academies (n 1).

[7] WTO, 'EU Proposal for WTO Disciplines and Commitments Relating to Electronic Commerce (INF/ECOM/22)' (2019) <https://trade.ec.europa.eu/doclib/docs/2019/may/tradoc_157880.pdf>.

when it comes to transparency and accountability of consumer-facing AI systems. When it comes to cross-border electronic commerce there is the question in how far an EU commitments under international trade law would have a bearing on EU policy on transparent and accountable AI. The WTO negotiations on trade-related aspects of electronic commerce are in full swing; however, the lack of transparency of these negotiations makes it difficult to oversee the current state of play.[8]

1. SCOPE, FOCUS AND METHODOLOGY

The scope of this study takes aim at constellations where EU law intersects with international trade law. Consider for a start the cross-border supply of digital services to EU consumers by businesses based outside the EU which is already quite commonplace in the global digital ecosystem. In addition, the digital services incorporate AI technology into their software architecture, as is the case already with many online services that use personalised online advertisements, content, recommendations, transactions and also prices. Two instances are considered: first, the supply of consumer-facing digital services from outside the EU; and second, the cross-border supply of AI technologies to EU organisations in the public and the private sector, to the extent that these technologies can affect EU consumers.

Outside the scope are AI technologies developed exclusively by public authorities and private actors based in the EU since they do not trigger EU trade law obligations in the first place.[9] Moreover, public services supplied in the exercise of governmental authority, such as law enforcement and public education, are not considered because they are exempted from the scope of the General Agreement on Trade in Services (GATS).[10] Finally, this study does not cover public procurement of AI technologies by the EU and member states[11] or how public procurement would fare under the WTO's Revised Agreement on Government Procurement.[12]

The focus is on consumer rights in global electronic commerce other than the right to protection of private life and personal data. The General Data Protection Regulation (GDPR) is only relevant for this analysis in so far that it regulates aspects of automated individual decision-making and profiling. The interface between EU's GDPR and digital trade deals has been the subject of a 2016 study which has helped to catalyse a change of EU trade policy to safeguard personal data protection.[13] However, individuals' right to the protection of privacy and personal data also conditions algorithmic transparency to a certain extent.

---

[8] See for an overview of the issues with transparency and for civil society representation Burcu Kilic and Renata Avila, 'Opening Spaces for Digital Rights Activism: Multilateral Trade Negotiations' (2020) <https://www.citizen.org/wp-content/uploads/Trade-Report_IPO-1.pdf>.

[9] The possible risks can, however, also occur in algorithmic and AI systems that have been developed in the EU. See for an overview of ADM systems in EU Member States AlgorithmWatch and Bertelsmann Stiftung, 'Automating Society: Taking Stock of Automated Decision-Making in the EU' (2019) <https://algorithmwatch.org/wp-content/uploads/2019/01/Automating_Society_Report_2019.pdf>; Fabio Chiusi and others, 'Automating Society Report 2020' (2020) <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/10/Automating-Society-Report-2020.pdf>.

[10] General Agreement on Trade in Services (GATS), Annex 1B to the 1994 Marrakesh Agreement on Establishing the World Trade Organization (WTO Agreement). GATS Article I 3 (c) defines "a service supplied in the exercise of governmental authority" as meaning "any service which is supplied neither on a commercial basis, nor in competition with one or more service suppliers." Examples are a Member's social security schemes or other public services, for example health or education, which are provided at non-market conditions.

[11] Government procurement has been effectively carved out, see GATS Article XIII.

[12] WTO, Revised Agreement on Government Procurement, Mar. 30, 2012, Annex 4(b) to the 1994 Marrakesh Agreement Establishing the World Trade Organization (WTO Agreement).

[13] Kristina Irion, Svetlana Iakovleva and Marija Bartl, *Trade and Privacy: Complicated Bedfellows? How to Achieve Data Protection-Proof Free Trade Agreements* (Institute for Information Law, University of Amsterdam 2016).

The methodology used in this study is legal and comparative research based on a comprehensive review of authorities and literature from the three legal domains covered. As much as possible, the study uses examples in order to demonstrate the interaction between AI governance and consumer rights in the EU and a new trade law discipline that protects source code the EU is willing to commit too.

2. STRUCTURE AND OUTLOOK

Section I recapitulates the current policies for holding AI-based decision-making systems accountable, and how this quest is supported by different transparency requirements. Both are a moving target, given that innovation and know-how on AI is progressing, as well as the knowledge and techniques to hold AI systems accountable. As a result, all that this section can achieve is to create a snapshot of our current understanding of the role of transparency for holding developers and providers of AI technology accountable and provide an outlook on future developments.

Section II introduces the spectrum of policy options currently discussed in connection with EU policy formulation on AI governance that is currently taking shape.[14] As well as the General Data Protection Regulation (GDPR),[15] which has rules on transparency, automated decision-making and profiling, an EU framework for trustworthy AI is currently being prepared. The Commission is planning to initiate legislation on safety, liability, and fundamental rights in the follow-up to its White Paper on Artificial Intelligence.[16]

Section III provides an overview over the anticipated risks of AI for consumers' rights in the Union. There are two sets of challenges for EU consumer protection: on the one hand, risks associated with AI-enabled consumer products and, on the other hand, risks for consumer rights in cross-border commerce. Both sets of consumer rights' challenges are likely exacerbated when AI's characteristic opacity (or 'black-box-effect')[17] obstructs oversight and enforcement of EU consumer protection law. This Section will demonstrate how crucial transparency, accountability and auditability of AI technology are for consumer rights and empowerment.

Next, Section IV covers EU's trade law obligations recognising that the GATS is presumed to cover cross-border trade in AI services.[18] Early in 2019, 76 WTO Members reinvigorated negotiations on trade-related aspects of electronic commerce to complement the GATS.[19] Attention is paid to the 'EU Proposal for WTO Disciplines and Commitments Relating to Electronic Commerce',[20] and a new discipline protecting source code of software, probing its substance in the light of the ongoing policy formulation for AI governance at the EU level.

Section V then combines the different strands of the argument in the preceding Sections and triangulates EU consumer protection standards with EU policy formulation in the

---

[14] Data Ethics Commission, 'Opinion of the Data Ethics Commission' (2019) <https://datenethikkommission.de/wp-content/uploads/DEK_Gutachten_engl_bf_200121.pdf>; European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)' (n 2).

[15] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016) OJ L 119, 4.5.2016, p. 1 [hereinafter GDPR] <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32016R0679>.

[16] European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)' (n 2).

[17] After the seminal book by Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2016).

[18] Kristina Irion and Josephine Williams, 'Prospective Policy Study on Artificial Intelligence and EU Trade Policy' (2020) 19f. <https://pure.uva.nl/ws/files/44322405/IViR_Artificial_Intelligence_and_EU_Trade_Policy.pdf>.

[19] WTO, 'Joint Statement on Electronic Commerce (WT/L/1056)' (2019) <https://docs.wto.org/dol2fe/Pages/FE_Search/DDFDocuments/251086/q/WT/L/1056.pdf>.

[20] WTO (n 7).

fields of AI, and with the EU position in the WTO electronic commerce negotiations. This Section by and large anticipates the compatibility of internal Union policies and rules in the field of consumer facing ADM systems. Additional consideration is given to preserving a margin of manoeuver to accommodate future developments as regards AI technology and associated risks inside EU governance instruments.

The Conclusions presents the findings of this study and makes nuanced policy recommendations that would help to improve the compatibility between internal Union policies while guarding a space of manoeuver for adapting requirements for transparent and accountable AI technology.

> **The central finding of this study is that such a source code clause being currently negotiated in plurilateral trade talks for a WTO agreement on electronic commerce, would restrict the EU's right to regulate in the field of AI governance in several important ways.**

I. TRANSPARENCY IN AI

The range of today's technologies which are referred to as artificial intelligence (AI) has an enormous potential for revolutionizing every aspect of contemporary society and individuals' lives. As a general purpose technology AI carries out functionalities across different sectors of social and economic life, and for this reason affects different types of users and interests. AI can improve the functioning of markets and public services but it can also disrupt various aspects of society. [21] The transformation from AI is expected to be more profound than that experienced with earlier general purpose technologies, such as the introduction of electricity.

It is the opacity of how AI learns and makes predictions that has captured our imagination.[22] Frequently, AI technologies are referred to as "black boxes" which is shorthand for the inscrutability of algorithmic decision-making.[23] This is how transparency has become valid currency in the ongoing debate on enabling accountability of and trust in AI. The concept of transparency, however, refers to a spectrum of different types of transparency; these in turn reflect the current state of knowledge about transparency motivated policy interventions with regards to AI. After clarifying some key concepts, this Section will summarise the ongoing debate on transparency in relation to AI supported decision making systems.

1. DEFINITIONS AND CONCEPTS

This Section starts with a brief outline of the frequently used technology-related concepts in EU policy documents. AI is used as an umbrella term for a variety of self-learning technologies, such as machine learning (ML). According to the European Commission, AI "refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals."[24] While reference to AI is catchy, it would be more correct to attribute current breakthroughs to the range of machine learning technologies which are the present state of the art.

a. Machine learning algorithms

For centuries the term algorithm has been used for an unambiguous mathematical formula. An algorithm consists of a set of rules that precisely define a sequence of operations to solve a specific problem.[25] A computer algorithm automates the calculation of such a set of rules. The leap forward is the scale and complexity with which contemporary computer algorithms process very large data sets. The empirical bedrock of most ML algorithms is applied statistics.

Today's ML technologies are the product of data analytics in which algorithms compute statistical probabilities from data sets provided by human programmers for training purposes. There are three branches of ML technologies:

1. when the data is structured and labelled the process is known as supervised learning;
2. when the data has not been labelled the process is called unsupervised learning; and

[21] Giovanni Sartor, 'Artificial Intelligence: Challenges for EU Citizens and Consumers (PE631.043)' (2020) 2 <http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/631043/IPOL_BRI(2019)631043_EN.pdf>.

[22] See Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data and Society 1 <https://doi.org/10.1177/2053951715622512>.

[23] Pasquale (n 17).

[24] European Commission, 'Artificial Intelligence for Europe (COM(2018) 237 Final)' (2018) <http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625>.

[25] Harold S Stone, *Introduction to Computer Organization and Data Structures* (McGraw-Hill 1972).

3. when supervised and unsupervised learning technologies are combined it is called reinforced learning.[26]

Based on the training data the ML algorithm is optimised for a particular goal and functionality, such as calculating insurance risks. A given ML algorithm is configured over numerous iterations of processing the training data until it produces the desired decision-making results with a reasonable level of accuracy. Put simply, the algorithm embodies decision-making rules that combine many different variables and their relative weights. Once this has been accomplished the ML algorithm can process new data.

A sub-group of ML technologies are self-learning which means that the algorithm continues to evolve the more data it has to make inferences from. In this case, the algorithm is not static but dynamic. A dynamic algorithm continues to optimize its underlying mathematical set of rules without instructions from human programmers. This means that even though the overarching goal and the initial parameters of the algorithm have been defined by its human developers, the operation and outputs of the ML algorithm continue to adapt to its environment.

b. Machine learning systems

A ML system, according to the European Commission, refers to "a collection of technologies that combine data, algorithms and computing power."[27] While this is broadly correct it disregards a few less prominent technical components and the non-technical properties of ML. Often a ML system is embedded in a larger software system to which it contributes, such as delivering personalised services. For a policy debate it does not suffice to exclusively focus on the technology in ML.

Research stresses that any AI/ML system is a socio-technical assemblage that combines and enacts human and non-human judgments.[28] This understanding indicates that resources, purpose, the choice and quality of training data, expertise and judgement, internal and external constraints have a significant influence on the technology.[29] A ML system should not be perceived "as a technical, objective, impartial form of knowledge or mode of operation"[30] but rather as a highly contextualised vehicle that serves an organisation's goals.[31]

> "The non-technical properties of these systems – for example, their purpose and constraining policies – are just as important, and often more important than their technical particulars."[32]

---

[26] The Royal Society, *Machine Learning: The Power and Promise of Computers That Learn by Example* (2017) 20.

[27] The White Paper incorporates by reference the more elaborate definition of an AI system given by the High Level Expert Group of Artificial Intelligence, see European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)' (n 2) 2.

[28] Rob Kitchin, 'Thinking Critically about and Researching Algorithms' (2017) 20 Information Communication and Society 14; Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media and Society 973, 974; AlgorithmWatch and Bertelsmann Stiftung (n 9); Mike Ananny, 'Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness' (2016) 41 Science Technology and Human Values 93.

[29] Kitchin (n 28).

[30] ibid.

[31] Brent Daniel Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society 1, 2 <http://journals.sagepub.com/doi/10.1177/2053951716679679>.

[32] Aaron Rieke, Miranda Bogen and David G Robinson, 'Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods' 1, 5 <http://omidyar.com/sites/default/files/file_archive/Public Scrutiny of Automated Decisions.pdf>.

c. Algorithmic decision-making

The term algorithmic decision making (ADM) is used to refer to the whole spectrum of algorithmic involvement in human decision-making.[33] On the one end of the spectrum, a decision can be algorithm-based but it is still an individual who makes the decision. On the other end of the spectrum, the decision is algorithm-determined to the extent that there is no intermediate step by any individual involved. In this situation it has been a human decision to delegate the execution of the decision to an ADM system.

ADM is not equivalent to the notion of automated individual decision-making, including profiling, as expressed in Article 22 of the GDPR. The GDPR notion of automated individual decision-making closely resembles an algorithm-determined decision from the ADM spectrum. However, in order for Article 22 of the GDPR to apply a decision based solely on automated processing must in addition produce a legal effect or significantly affect an individual. Where this is the case, the GDPR provides for the right to obtain human intervention, the right to be heard and to contest the automated decision.

There are good arguments why the whole spectrum of ADM should be addressed by regulation instead of focusing only on fully automated decision-making. It is contested whether a human in the loop would make a difference for the quality of the ADM. Research has found that humans tend to fall in line with computational judgements, a tendency which is known as automation bias.[34] In order to become a meaningful protection mechanism human intervention needs to be carefully designed and provide incentives to contest algorithmic judgements.

d. Probability rather than knowledge

Even though the methods are borrowed from the exact sciences a ML algorithm produces associations that are short of cognition. The training data is unlikely to convey a true model of reality and neither can an algorithm deliver causal explanations for its conclusions.[35] It is not knowledge but probability that informs the decision-making rules of algorithms:

> "Much algorithmic decision-making and data mining relies on inductive knowledge and correlations identified within a dataset. Causality is not established prior to acting upon the evidence produced by the algorithm."[36]

ML algorithms produce predictive systems which are often void of an explanation for their decisions, but are nevertheless put to commercial use.[37] This makes ADM systems vulnerable to either reproduce bias from the training data, or to learn to unfairly discriminate between individuals. Consider in addition that the complexity of many ML algorithms can exceed human capabilities, including that of the developers of the very ADM system. The resulting inscrutability in turn increases the risk that mistakes and biases of a given ADM system can go undetected for quite some time.

Where a corporate culture of 'move fast and break things' prevails, commercial strategies tend to prioritize time to market over careful prototype testing for unintended effects of the technology for individuals' rights and interests. Here public policy and regulation has an important role to play in providing incentives for rigorous impact assessments and testing before introducing a new ADM system that can affect individuals and society.

---

[33] Data Ethics Commission (n 3) 161.

[34] ibid 163.

[35] Dana Mackenzie and Judea Pearl, *The Book of Why: The New Science of Cause and Effect* (Basic Books 2018).

[36] Mittelstadt and others (n 31) 5.

[37] Jonathan Zittrain, 'The Hidden Costs of Automated Thinking' *The New Yorker* (New York, 2019) <https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking> accessed 30 July 2019.

Developers and providers have to resume full accountability for the predictions and outcomes of the algorithms they implement as well as legally responsible for its lawful operations.

e. Scalability of machine learning systems

Scalability is a feature of ML technology which allows it to compute as many operations simultaneously as required provided there is sufficient computing power. In order to grasp the dimensions consider for example that the AI system 'AlphaGo Zero' played over 4.9 million games of the Chinese board game Go against itself over the period of three days.[38] This has been the training that this ML system required to achieve superhuman performance in the rules of this particular game and win against earlier versions of itself.

The sheer capacity of a fully operational ML system to serve instantaneously a very high number of its users is difficult to grasp. There are already ML systems today that operate across an entire population of users, such as the targeting of online advertisements or the personalised recommendations systems of social networks. AI's ability to perform at a massive scale can translate into a competitive advantage for innovating firms. An AI system that is successful in performing a particular task, could rapidly take hold and transform an entire sector of human activity.

In the policy debate, however, AI's ability to perform at scale has not yet reached the prominence it deserves. Academics point out that:

> "… the speed and scale at which these technologies now operate poses novel threats, risks and challenges which contemporary societies have not hitherto had to contend with."[39]

In particular, the question how the law and regulatory tools can deal with the scalability of ADM systems is still in its infancy. On the one hand, it should be noted that private and public enforcement are not scalable in the same way as ADM systems are. On the other hand, legal protections are mostly geared towards individual rights as opposed to collective interests and societal values:

> "Because current approaches to the interpretation and enforcement of human rights are highly individualized in orientation, they are likely to struggle to address the collective, aggregate and cumulative risks and harms that these technologies might generate."[40]

We will return to this question in Section III when asking how fit the regulatory oversight and redress mechanisms in consumer protection are to deal with population-wide ADM systems.

f. Cross-border effects of ADM systems

From the outset the Commission recognizes that AI is easily tradeable across borders.[41] In the case of AI-enabled products or services it is not uncommon that developers, vendors, customers and users of an algorithmic system are spread around the world:

> "We are coming into a world in which your credit, your job prospects, your insurance claim, the news you read, and even the dates you go on are determined by faceless computers in a distant land."[42]

---

[38] David Silver and others, 'Mastering the Game of Go without Human Knowledge' (2017) 550 Nature 354, 255f.

[39] Karen Yeung, 'Responsibility and AI' (2019) 42 <https://rm.coe.int/responsability-and-ai-en/168097d9c5>.

[40] Ibid.

[41] European Commission, 'Artificial Intelligence for Europe (COM(2018) 237 Final)' (n 24).

[42] Anupar Chander, 'AI and Trade' in Mira Burri (ed), *Big Data and Global Trade Law* (Cambridge University Press 2020).

The reason is that many ADM systems are integrated into the global data-processing infrastructure and their predictive outcomes can be applied across today's digital ecosystem.[43]

> "It is important to recognise that the global interconnectivity and reach of the internet (and internet-connected technologies) have enabled the swift roll-out of AI technologies on a massive scale, particularly with the rapid and widespread take-up of 'smart' networked devices."[44]

While the transnational provision of AI has many benefits it also increases the interdependence between different domestic legal frameworks.[45] Societies, by contrast, have diverse set-ups of rights, freedoms and legal protection mechanism which do not necessarily converge. Without due consideration for the local societies and legal frameworks of the receiving countries and regions, the cross-border supply of ADM does not only risk to undermine legal safeguards but it can also be a poor fit for the local circumstances and users.

When an AI system is applied across borders it can have repercussions for the societies it interacts with, both at individuals and societal levels. These repercussions can result from regulatory differences between countries, where a certain practice is unregulated in the country of origin but regulated or even prohibited in the receiving country. Take an automated marketing tool, for example, that violates EU law on unfair commercial practices but is permissible in the law of the country where the provider is established. The cross-border supply of AI products and services can counteract legal protection mechanisms across the board, such as the protection of personal data, consumer rights, and anti-discrimination law.

2. A PRIMER ON ALGORITHMIC TRANSPARENCY

In light of the challenges, as outlined above, to ensure that developers and providers of AI technology comply with domestic laws, the quest for transparency is strongly justified. The ideal of algorithmic transparency follows the logic that "observation produces insights which create the knowledge required to govern and hold systems accountable."[46] However, transparency is not a magic wand but needs to be carefully managed in order to benefit algorithmic governance in the EU.[47] Focusing transparency exclusively on certain technical components of an AI system, such as the algorithm or the training data, is not the same as holding all its technological and social aspects accountable.[48] Following an explanation of some key concepts, this Section will map out the range of transparency instruments and explain how they are currently combined and applied.

a. Proprietary and open source code

The source code of ML algorithms can be proprietary or open source. Proprietary (i.e. privately owned) algorithms are generally not open for inspection since they contain business secrets.[49] Known examples of proprietary secret algorithms include Google's search PageRank algorithm or its Maps navigation algorithm. Preserving the competitive

[43] Irion and Williams (n 18) 3; Giovanni Sartor, 'Artificial Intelligence: Challenges for EU Citizens and Consumers' (European Parliament 2019) 3.

[44] Yeung (n 39) 22.

[45] UN Secretary-General's High-level Panel on Digital Cooperation, 'The Age of Digital Interdependence: Report of the UN Secretary-General's High-Level Panel on Digital Cooperation' (2019) <https://digitalcooperation.org/wp-content/uploads/2019/06/DigitalCooperation-report-for-web.pdf>.

[46] Ananny and Crawford (n 28) 974.

[47] ibid.

[48] Ananny and Crawford (n 28).

[49] Mittelstadt and others (n 31) 6.

advantage of a well-functioning algorithm is the frequent justification for keeping algorithms secret.[50] Another frequent justification in support of an algorithm's secrecy is that, by having this information, the algorithm could be manipulated.[51] A proprietary algorithm can even qualify for protection as a trade secret under EU law.[52]

There is a good range of open source (i.e. publicly available) libraries that provide ready-to-use algorithms to perform standard operations in ML, such as regression analysis or clustering data. For a number of standard tasks, such as image recognition and natural language processing, there are even pre-trained algorithms available to developers. This indicates that training data is often more important than the ML algorithms that is used to compute it. Even when ML incorporates an open source code, this is re-configured based on the data it is trained on.[53] The properties of the resulting trained algorithm are then often no longer in the public domain, but become proprietary. This ultimately makes open source algorithms the exception rather than the rule.

b. White box and black box testing

Audits which enable the assessment of algorithms, their data and design processes, can discover illegal or unethical practices and consequently can hold developers and providers accountable.[54] Policy and research documents commonly distinguish two mechanisms for testing and auditing how a ML system functions: "white box" method and "black box" method.

White box testing is a method to audit an algorithm which involves an analysis of its source code.[55] White box testing can be used in internal audits and in situations where the source code is accessible. Because white box testing requires resources, specialized knowledge and access to proprietary information this method should be used in a targeted investigation to identify the source of an existing concrete problem.[56] For example, this type of code review has been successfully used by U.S. academics to identify the offending section of code in the U.S. regulatory investigation of the Volkswagen Diesel nitrogen oxide emission cheating scandal.[57]

In contrast, black box testing develops all the techniques used to interrogate the workings of a ML algorithm, without the need to access its source code. The most accessible of these techniques is based on observing the inputs and outputs of an algorithmic system, which can then be used to experiment with public-facing online services, such as online recommendation systems.[58] There are more sophisticated techniques for black box testing, however, they quickly regain the character of scientific investigations which require

---

[50] Mariateresa Maggiolino, 'EU Trade Secrets Law and Algorithmic Transparency' [2019] SSRN Electronic Journal 1, 1.

[51] Nicholas Diakopoulos, 'Accountability in Algorithmic Decision Making' (2016) 59 Communications of the ACM 56, 62; Mittelstadt and others (n 31) 6.

[52] Maggiolino (n 50).

[53] Mario Martini, 'Fundamentals of a Regulatory System for Algorithm-Based Processes. Expert Opinion Prepared on Behalf of the Federation of German Consumer Organisations (Verbraucherzentrale Bundesverband)' (2019) 7 <https://www.vzbv.de/sites/default/files/downloads/2019/07/19/martini_regulatory_system_algorithm_based_processes.pdf>.

[54] High-Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI' (2019) 19 <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419>; Sandvig and others (n 26); Bodo and others (n 57).

[55] Rieke, Bogen and Robinson (n 32) 19.

[56] ibid; Ansgar Koene and others, 'A Governance Framework for Algorithmic Accountability and Transparency' (2019) <http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf>.

[57] See Megan Geuss, 'A Year of Digging through Code Yields "Smoking Gun" on VW, Fiat Diesel Cheats' *Ars Technica* (28 May 2017) <https://arstechnica.com/cars/2017/05/volkswagen-bosch-fiat-diesel-emissions-cheats-cracked-open-in-new-research/>; Rieke, Bogen and Robinson (n 32) 19.

[58] Rieke, Bogen and Robinson (n 32) 17.

expertise and resources for the empirical investigations and statistical analysis.[59] Applying black box testing can:

> "allow examiners to draw reliable, sophisticated conclusions about how an automated system functions even without access to the system's source code."[60]

c. Interface audits

Though access to proprietary algorithms is not necessary, in order to carry out more sophisticated black box testing experts must gain access to input and output data. In the case of proprietary AI systems both access to the input and output data and to the respective interfaces is not always publicly possible. In many situations developers and providers treat them just as proprietary as the algorithms behind them.

The quest for observability oftentimes concentrates on gaining access to the interfaces of an AI system (so called Application Programming Interfaces (API)), which provide the gateways through which the algorithm receives its inputs and produces some sort of output.[61] It is not a coincidence that researchers discuss the potential of requiring access to interfaces of AI systems for accountability purposes across several legal domains, such as competition law, anti-discrimination law, online platform regulation and broadly AI governance.[62]

In practice each algorithmic system's architecture is unique which "makes interacting with it programmatically much less standardized."[63] Mandating standardized auditing interfaces may solve the problem with accessing the necessary data:

> "Likewise, operators should be obliged to use adequate and interoperable IT solutions when implementing interfaces to enable official introspection."[64]

> "However, in areas with a high potential for harm, it may be necessary to stipulate that system operators must use a standardised interface."[65]

In the context of AI applications with serious potential for harm, such as an algorithm to determine "the creditworthiness of an individual consumer or company",[66] the Data Ethics Commission recommends to enable "always-on" oversight via a live interface with the

---

[59] Sandvig and others (n 26) describe five auditing designs: (1) code audit, (2) noninvasive user audit, (3) scraping audit, (4) sock puppet audit, and (5) collaborative or crowdsourced audit; see also Carsten Orwat, 'Risks of Discrimination through the Use of Algorithms' (2020) 70 <https://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/EN/publikationen/Studie_en_Diskriminierungsrisiken_durch_Verwendung_von_Algorithmen.pdf?__blob=publicationFile&v=2>; Bodo Bodo and others, 'Tackling the Algorithmic Control Crisis – the Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents' (2017) 19 The Yale Journal of Law & Technology 133.

[60] Rieke, Bogen and Robinson (n 32) 18.

[61] Martini (n 49) 28; Rieke, Bogen and Robinson (n 32) 11; Sandvig and others (n 54).

[62] Bernhard Rieder and Jeanette Hofmann, 'Towards Platform Observability' (2020) 9; Martini (n 53); Koene and others (n 56); Guillaume Klossa, 'Towards European Media Sovereignty: An Industrial Media Strategy to Leverage Data, Algorithms and Artificial Intelligence' (2019) 60; Ian Brown, 'The Technical Components of Interoperability as a Tool for Competition Regulation' (2020) <https://openforumeurope.org/wp-content/uploads/2020/11/Ian_Brown_The_technical_components_of_interoperability_as_a_tool_for_competition_regulation.pdf>.

[63] Sandvig and others (n 54) 8.

[64] Martini (n 53) 28.

[65] Data Ethics Commission (n 3) 200.

[66] ibid 180.

algorithm.[67] Another concrete proposal calls for the introduction of 'accountability interfaces'[68] to ensure the observability of algorithmic systems through providing access to a continuous stream of data to and from the algorithmic system.

In addition, interface audits of an algorithms can be carried out in a sandbox (i.e. an isolated testing environment), where data inputs come from consenting research participants (similar to panel research and surveys today) or synthetic users (so-called sock puppet audits).[69] The proposal to harness transparency obligations to expose specific APIs in order to create algorithmic sandboxes has gain traction in relation to digital media platforms.[70] The upshot is that interfaces (APIs) of AI systems are bound to become important gateways for algorithmic transparency.

d. Balancing business secrets, data protection and transparency

From what has been said above it emerges that with many proprietary systems there is a steep information asymmetry between the developers and providers of AI systems and its users and the public at large. If algorithmic systems are entirely shielded as business secrets from external inspection, meaningful control and oversight would simply not be feasible.[71] Future transparency rules thus have to strike a balance between the protection of the business interests and the public interest to hold AI systems legally accountable.

This is where prospective transparency laws will come into play in order to enable a good measure of checks and balances over ADM systems. The least controversial today are proposals to legislate some form of qualified transparency which would authorize regulatory authorities and domestic courts to request access to data, interfaces and algorithms for investigations that would selectively override business secrets of developers and providers.[72]

In addition, it is important to note that the different methods to scrutinize algorithmic systems can have implications for the protection of privacy and personal data of individual users. Introducing transparency at the level of an algorithm's source code would often not require access to individuals' personal data.[73] However, leveraging source code transparency tends to be more invasive to an operator's business secrets. Conversely, conducting research into algorithms by means of gathering input and output data will be often the best method to gain insights into how the algorithmic system operates but it is potentially data privacy invasive, if personal data of individual users is exposed.

Ensuring interfaces audits in justified situations can be a good compromise that respects the algorithm's business secret and users' data privacy. Providing access to an algorithm's interfaces could be a meaningful transparency obligation in the context of many consumer-facing AI services that would enable regulatory authorities and other public interest organisations to observe an algorithm using non-privacy invasive methods.

3. MAPPING THE RANGE OF TRANSPARENCY INSTRUMENTS

There is no one-size fits all solution to what is the optimal level of algorithmic transparency but optimal levels of transparency must be situated within contexts of technologies, practices and social domains. In fact several transparency mechanisms can co-exist parallel to each other and transparency obligations should escalate in situations that warrant a closer look under hood of a given AI system or even a sector-wide enquiry. Thus, a

[67] ibid 179.

[68] Rieder and Hofmann (n 62) 15.

[69] Koene and others (n 56) II; Sandvig and others (n 54); Rieder and Hofmann (n 62) 18.

[70] Klossa (n 62) 60.

[71] Koene and others (n 56).

[72] ibid 51.

[73] Data Ethics Commission (n 3) 170.

modular approach to algorithmic transparency will be needed which combines information duties for users and consumers with public scrutiny and investigatory powers by public authorities and judicial review.

Algorithmic transparency is best understood as a multi-dimensional concept. Different configurations emerge from combining relevant elements from the following three dimensions: substantive, personal and temporal. The substantive dimension distinguishes whether transparency is called for obtaining a general description of the functioning of a given ADM system, or serves to allow an external inspection in the form of an audit to evaluate a given ADM system. Different intervention points can be necessary to facilitate an external inspection, such as access to the source code of the algorithm, interfaces, audit logs, training datasets and so forth.

The personal dimension defines the intended audience of any transparency requirement which can be the individual users of an AI system, the general public, external auditors, supervisory authorities or domestic courts. Transparency requirements must be carefully designed to take the different levels of expertise of the intended audience into account ranging from laymen to expert audiences. For example, the disclosure of the source code to external auditors and supervisory authorities may help them to inspect the system provided they have the necessary technical expertise. Ordinary users and consumers by contrast typically cannot comprehend highly technical explanations about the functioning of an ADM system.[74] They need easily comprehensible information about the algorithmic system and an explanation how the decision concerning them came about and which factors had what influence.

The temporal dimension of transparency determines whether transparency of the functioning of an algorithmic system is called for before, during or after its use and it may in exceptional circumstances even involves continuous oversight. The presence of dynamic ML algorithms and correspondingly adaptive AI systems further complicates the timing of transparency:

> "This is particularly true for adaptive systems that 'learn' as the amount and types of data they draw on increase—and for platforms with shifting interfaces, settings, capabilities, and number of users. There is no 'single' system to see inside when the system itself is distributed among and embedded within environments that define its operation."[75]

Table 1 below provides an overview over the available options currently discussed in relation to affording transparency about algorithmic and AI supported decision making systems.

A MODULAR APPROACH TO ALGORITHMIC TRANSPARENCY IS NEEDED

Algorithmic transparency does not only provide actionable information to affected individuals and qualified transparency for public authorities and courts. In justified circumstances, transparency is also called for in order to effectuate collective redress and public interest research. Algorithmic audits are central to accountability, verifiability, and trust in AI. Currently, input/output audits ("black box" method) are used more frequently than auditing an ML algorithm's source code ("white box" method). Interfaces (APIs) of AI systems are bound to become important gateways for algorithmic transparency.

[74] Martini (n 53) 12; Sandra Wachter and others, 'Counterfactual Explanations Without Opening the Black Box: Automated Decsions and the GDPR' (2018) 31 Harvard Journal of Law & Technology 842f.

[75] Ananny and Crawford (n 28).

Table 1 Overview of transparency regimes for algorithmic decision making systems

	Intervention	Purpose	Level of engagement	Timing	Target audience	Status and issues
Lay audience	Identification	Signal the existence of algorithmic decision making (ADM)	Descriptive general	Ex ante and ex post	General public	Partially regulated in GDPR No public information requirement yet First AI registry in the public sector launched[76]
	Information	Information about logic, data used and scope of algorithm-based processes	Descriptive general	Ex ante	Individuals, users and consumers	Partially regulated in GDPR
	Explanation	Comprehensible explanation of the result of a decision	Descriptive specific	Ex post	Affected individuals, users and consumers	Partially regulated in GDPR Academics promote concept of counterfactual explanation[77]
General Public	Freedom of information	Access to government-held records	Policies, procurement documents, access to source code	Ex ante and ex post	External, subject to freedom of information request	So far only in French law[78]

[76] The City of Amsterdam, Algorithm Register, 30 September 2000 <https://algoritmeregister.amsterdam.nl/en/ai-register/> accessed 5 November 20

[77] Wachter and others (n 74).

[78] The Digital Republic Bill (*Loi pour une République Numérique*, n° 2016-1321) of 7 October 2016, Article 2(I), (in French) <https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte>.

	Risk impact assessment	Publication of documentation	Descriptive general	Ex ante	General public information	No legal obligation to publish impact assessment[79]
Public scrutiny	Public interface audits[80]	Experiment with algorithmic systems	Public-facing interface, input/ output audit (blackbox methods)	Ex post	Public interest research, experts	Difficulties to access proprietary interfaces (APIs) Web-scraping often not permitted by operators The collection of user data must respect the GDPR
Experts	External audits	External assessment and compliance check	Input/ output audit, access to data, interfaces (black-box methods) Benchmarking:[81] Disclosure of system's key statistics, error rates, accuracy level (or confidence values) Access to source code, training data, decision models, audit logs, interfaces, confidence values (white box methods)	Ex ante and ex post	(Independent) external auditor	External audits not yet mandated by law, only in sector-specific laws such as for medical device. External audits can be contractually stipulated, e.g. in public procurement contracts[82]

[79] Risk impact assessment are recommended by Data Ethics Commission (n 3) 188.

[80] Sandvig and others (n 54); Jef Ausloos, Paddy Leerssen and Pim ten Thije, 'Operationalizing Research Access in Platform Governance What to Learn from Other Industries?' (2020) <https://algorithmwatch.org/wp-content/uploads/2020/06/GoverningPlatforms_IViR_study_June2020-AlgorithmWatch-2020-06-24.pdf>; Bodo and others (n 54).

[81] Diakopoulos (n 51).

[82] See AI Now Institute, City of Amsterdam, City of Helsinki, Mozilla Foundation and Nesta, 'Using procurement instruments to ensure trustworthy AI', 15 June 2020, <https://assets.mofoprod.net/network/documents/Using_procurement_instruments_to_ensure_trustworthy_AI.pdf> accessed 5 November 2020; City of Amsterdam, 'Standard Clauses for Municipalities for Fair Use of Algorithmic Systems', 2020 <https://www.amsterdam.nl/wonen-leefomgeving/innovatie/de-digitale-stad/grip-op-algoritmes/> accessed 5 November 2020.

Regulatory authorites	Authorization	Market entry requirement	Input/ output audit, access to data, interfaces (“black-box” methods) Access to source code, training data, decision models, audit logs, interfaces, confidence values (white box methods)	Ex ante	Supervisory authority or appointed external body	Currently not legally mandated Recommended for high risk AI systems[83]
	Regulatory enforcement	Compliance check against legal requirements Ex officio or based on an individual’s complaint	Input/ output audit, access to data, interfaces (black-box methods) Access to source code, training data, decision models, audit logs, interfaces, confidence values (white box methods)	Ex post	Competent supervisory authority (anti-discrimination bodies, consumer protection authorities, data protection authorities, competition law authorities, others) Possibly delegation to external auditor	Legal competence and powers to carry out investigations in the form of audits, currently recognised in the GDPR and EU competition law for example
Domestic courts	Judicial review	Private or collective redress against an algorithmic or automated decision	Any method mentioned above as the Court sees fit	Ex post	Court appointed external auditor or IT forensic expert	Initial burden of proof on the claimant Collective redress mechanisms in the GDPR and expected for consumer protection[84]

[83] The Opinion of the Data Ethics Commission advices prior authorization of AI systems that are deemed high-risk for affected individuals, groups or the society at large. Data Ethics Commission (n 3) 179.

[84] See Article 80 of the GDPR; new legislating pending European Parliament and the Council, Directive on representative actions for the protection of the collective interests of consumers, and repealing Directive 2009/22/EC (unpublished final text) <https://www.consilium.europa.eu/media/44766/st09223-en20.pdf>.

II. EMERGING AI GOVERNANCE

Artificial intelligence has become a major policy issue for the EU and its member states. The Commission consistently emphasizes its vision to advance AI on the basis of the Union's values as stipulated in Article 2 of the Treaty on European Union (TEU) and in line with the Charter of Fundamental Rights of the EU. In its 2018 Communication 'Artificial Intelligence for Europe' the Commission announces three major lines of action:

1. Promote research, development and uptake of AI;
2. Prepare for socio-economic changes brought about by AI; and
3. Ensure an appropriate ethical and legal framework.[85]

With the exception of privacy and personal data protection, the tenets of what makes AI responsible are not (yet) codified in EU law. This Section will focus on the ongoing development of an ethical and legal framework covering AI technology, taking recourse to two of the current proposals for its transparency and accountability. In line with the terms of reference, this study will primarily draw on the Opinion of the German Data Ethics Commission and the Commission's White Paper on Artificial Intelligence.[86]

1. THE OPINION OF THE DATA ETHICS COMMISSION

Next to many noteworthy initiatives on ethical and trustworthy AI, this Section will summarize the influential 2019 Opinion by the German Data Ethics Commission.[87] The Opinion calls on the EU legislator to adopt "a risk-adapted regulatory approach" to algorithmic systems that distinguishes between five levels of criticality.[88] This approach incorporates the principle that "the greater the potential of algorithmic systems to cause harm, the more stringent the requirements and the more far-reaching the intervention by means of regulatory instruments."[89]

a. Recommendations for a risk-adapted regulatory approach

The Opinion proposes the concept of a criticality pyramid with five levels:

- Level 5 Applications with an untenable potential for harm
- Level 4 Applications with serious potential for harm
- Level 3 Applications with regular or significant potential for harm
- Level 2 Applications with some potential for harm
- Level 1 Applications with zero or negligible potential for harm

The threshold for regulatory intervention would start with application as of level 2, progress in intensity for levels 3 and 4 until the grave risks at level 5 would command a complete or partial ban. Regulation designed for level 2 applications would primarily rely

[85] European Commission, 'Artificial Intelligence for Europe (COM(2018) 237 Final)' (n 24).

[86] European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)' (n 2); Data Ethics Commission (n 3).

[87] Data Ethics Commission (n 3).

[88] ibid 173f. *See* Tobias Krafft / Katharina Zweig, 'Transparenz und Nachvollziehbarkeit algorithmenbasierter Entscheidungsprozesse' [Transparency and traceability of algorithm-based decision processes], Study commissioned by the Federation of German Consumer Organisations (vzbv) (2019) <https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-01-22_zweig_krafft_transparenz_adm-neu.pdf>.

[89] ibid 173.

on ex-post controls and enforcement. As of level 3 applications would need an ex-ante approval before being introduced, and facilities for continuous supervision at level 4.[90]

b. Recommendations on the regulatory architecture

The Opinion recommends the introduction of horizontal requirements for algorithmic systems as an EU regulation which are to be supplemented with sectoral instruments.[91] The envisioned EU horizontal regulation should provide for general substantive rules "on the admissibility and design of algorithmic systems, transparency, the rights of individuals affected, organisational and technical safeguards and supervisory institutions and structures."[92] This construction would ensure that there is a unified level of regulation throughout the EU "which sets out basic principles for all algorithmic systems"[93], thereby contributing to legal certainty for operators of algorithmic systems in the public and the private sector and EU citizens alike.

The Data Ethics Commission argues that sectoral instruments offer a way for targeting regulatory intervention without overburdening a future EU horizontal regulation. The sectoral instruments are meant to supplement the horizontal regulation "with specific provisions for individual sectors or potentially harmful situations".[94] This holds the advantage to allow for desirable "differentiation between the different needs for protection involved for individual systems and usage contexts"[95] that would moreover recognize the distribution of competences between the EU and member states.

c. Recommendations for basic regulatory principles at EU level

The Data Ethics Commission recommends to introduce as basic principles:

- a mandatory labelling scheme for algorithmic systems;
- affected individuals' should have a right
    - to meaningful information about the logic involved, as well as the scope and intended consequences of an algorithmic system;
    - In certain situations a right to provide an individual explanation of a decision;
- a conditional right to information for journalistic and research purposes and an unconditional right of access to information on algorithmic systems with serious potential for harm (Level 4) used by the state;
- an obligation for operators of regulated algorithmic systems to produce and publish a risk assessment covering:
    - potential risks for self-determination, privacy, bodily integrity, personal integrity, assets, ownership and discrimination;
    - information on the underlying data and logic of the model, methods for gauging the quality and fairness of the data and the model accuracy;
- further requirements to document and log the data sets and models used, the level of granularity, the retention periods and the intended purposes intended for supervision and enforcement;

[90] Namely an option for "always-on" regulatory oversight via a live interface with the algorithmic system, ibid 179.

[91] ibid 180.

[92] ibid 280.

[93] ibid 180.

[94] ibid 181.

[95] ibid 180.

- additional protective mechanisms for all algorithmic decision making, irrespective of whether they are algorithm-supported, -based or determined;
- licensing procedures or preliminary checks of algorithmic systems with regular or significant (Level 3) and serious potential for harm (Level 4),
- additional protections against discrimination by algorithms complementing existing anti-discrimination laws.[96]

d. Recommendations for governance and enforcement

In its opinion, the Data Ethics Commission gives due attention to oversight mechanisms and institutions as well as cooperation between competent national authorities of the member states and EU bodies.[97] One pertinent issue is the question of building and providing the specialized expertise that would be necessary to carry out supervision and enforcement activities by competent authorities. The Opinion recommends the set-up of competence centres for algorithmic systems at member states and EU levels which function as a knowledge hub and support competent supervisory authorities.[98]

A rather noteworthy recommendation is the proposal to facilitate “always-on” regulatory oversight of algorithmic systems which exhibit a high potential for harm (Level 4) through a live interface with the system.[99] Standardised interfaces could be used “to carry out what are known as input-output tests, which check, for example, whether an algorithmic system systematically discriminates against groups.”[100] This is an innovative proposition that would require a mandatory requirement for operators of covered algorithmic systems to set up and provide access via such standardized interfaces.

2. WHITE PAPER ON ARTIFICIAL INTELLIGENCE

The Commission has the power of initiative in EU law-making (Article 17(2) TEU). Early 2020, the Commission published a ‘White Paper on Artificial Intelligence’[101] which sets out policy options for prospective AI regulation and governance in the EU. In the White Paper the Commission promotes an EU-wide approach in order to prevent the fragmentation of rules pertaining to AI in the internal market and sets forth the options for future decision-making in this domain.

The White Paper envisions an “ecosystem of trust”[102] in which consumers and businesses in AI can rely on a clear European regulatory framework. It is not that AI currently operates in a legal vacuum but there are a number of legislative instruments that would already govern activities involving AI. For instance, the GDPR is a piece of legislation that applies to the processing of individuals’ personal data and governs automated individual decision-making, including profiling. Moreover, consumer protection law is already in place even if certain rules, such as product liability, would need upgrading in light of technical developments.

---

[96] ibid 196f.

[97] ibid 198.

[98] ibid 199.

[99] ibid 184.

[100] ibid 199.

[101] European Commission, ‘White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)’ (n 2).

[102] ibid 3.

a. Thresholds for regulatory intervention

When it comes to designing mandatory legal requirements for AI applications the White Paper projects a risk-based approach to future AI regulation to ensure that any "regulatory intervention is proportionate."[103] The Commission argues that the threshold for regulatory intervention should be "clear and easily understandable." The Commission envisions to pass new regulation for AI applications which are deemed 'high-risk'.

Following the White Paper's proposals the determination whether a given AI application is deemed 'high-risk' would require that two cumulative criteria are met: "both the sector and the intended use involve significant risks."[104] The 'high-risk' sectors that constitute the first criterium should be "specifically and exhaustively listed in the new regulatory framework." The second criterium that considers the "intended use" would be used to determine which AI applications from the listed sectors are deemed 'high-risk' following their specific functionality and use. Under the second criterium mundane AI applications operating in the 'high-risk' sectors would be sorted out while application that produce significant risks, "in particular from the viewpoint of protection of safety, consumer rights and fundamental rights",[105] would be regulated.

As a separate category, the Commission recognizes that certain AI application should be considered 'high-risk' per se, irrespective of the sector in which they operate. As illustrations the White Paper highlights the use of AI applications for recruitment processes or remote biometric identification, such as facial recognition technology.[106]

b. Mandatory requirements of AI governance

When it comes to the substance of future AI regulation, the White Paper takes recourse to the "Ethics Guidelines for Trustworthy AI" of the High Level Expert Group on AI.[107] Future regulation addressing high-risk AI applications would stipulate mandatory requirements relating to training data, data and record-keeping, information duties, robustness and accuracy as well as human oversight.[108] For training data, among others, the White Paper envisages to place "obligations to use data sets that are sufficiently representative"[109] across "relevant dimensions of gender, ethnicity and other possible grounds of prohibited discrimination."[110]

The White Paper sets out extensive record-keeping duties "in relation to the programming of the algorithm, the data used to train high-risk AI systems, and, in certain cases, the keeping of the data themselves."[111] With respect to the algorithm the proposal would require "documentation of the programming and training methodologies, processes and techniques used to build, test and validate the AI system."[112] There is, however, no requirement foreseen to keep a record of the actual algorithm which may change over time or (auditing) logs of the AI system.

---

[103] ibid 17.

[104] ibid.

[105] ibid.

[106] ibid 18.

[107] High-Level Expert Group on AI (n 54).

[108] European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)' (n 2) 18.

[109] ibid 19.

[110] ibid.

[111] ibid.

[112] ibid.

The intention for such data and record-keeping as well as documentation is to allow for retroactive control and verification. “This should not only facilitate supervision and enforcement”, the White Paper notes, “it may also increase the incentives for the economic operators concerned to take account at an early stage of the need to respect those rules.”[113]

The envisaged information duties are descriptive information about the AI system’s capabilities and limitations for deployers of the system, competent authorities and affected third parties. Citizens should be aware that they are interacting with an AI system and have access to objective, concise and easily understandable information about the AI system.

According to the White Paper, ensuring robustness and accuracy should be requirements placed on high-risk AI applications.[114] The benchmarks for ‘high-risk’ AI systems are that they must be robust and accurate corresponding to their level of accuracy and outcomes reproducible. Moreover, AI systems should be able to adequately deal with errors or inconsistencies, and resilient against overt attacks and attempts to manipulate the algorithm or the data.

A right to human oversight is also discussed as a potential safeguard in future EU regulation of AI.[115] The envisaged requirement of human oversight is primarily geared towards not being subject to ADM, to request human review and the ability for a human to overrule an AI system.[116]

c. Governance and enforcement

The White Paper sets out a two-tiered governance structure consisting of ex ante conformity assessment and ex post supervision and enforcement.

The conformity assessment should take the form of an independent audit and assessment of whether an AI-system complies with the mandatory requirements of a prospective AI regulation. The prior conformity assessment could “include checks of the algorithms and of the data sets used in the development phase.”[117] The rational for introducing ex ante independent conformity assessments is to increase trust and ensure objectivity.[118] The Commission proposes to entrust conformity assessments to notified testing centres designated by member states thereby building on an existing EU system of conformity assessment for products designated for the EU internal market and the development of a European cybersecurity certification scheme pertaining to ICT products, services and processes.[119]

The ex-ante conformity assessment should be “without prejudice to monitoring compliance and ex-post enforcement by competent national authorities.”[120] The Commission’s

---

[113] ibid.

[114] ibid 20.

[115] ibid 21.

[116] ibid.

[117] ibid 23.

[118] ibid 25.

[119] *See* for conformity assessment of products Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC, OJ L 218, 13.8.2008, p. 82–128. *See* for the development of a European cybersecurity certification scheme pertaining to ICT products, services and processes Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), OJ L 151, 7.6.2019, p. 15–69.

[120] European Commission, ‘White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)’ (n 2) 24.

White Paper correctly notes that “some specific features of AI (e.g. opacity) can make the application and enforcement of this legislation more difficult.”[121] Ex-post controls can be conducted on the one hand based on the required documentation and on the other hand involve testing regulated high-risk AI applications. According to the White Paper it would be for competent authorities to request the records and documentation and where relevant data sets for testing and inspection. “Where necessary, arrangements should be made to ensure that confidential information, such as trade secrets, is protected.”[122]

d. Cross-border supply of AI

In its White Paper the Commission demonstrates great awareness of the cross-border supply of AI. Not unlike the GDPR’s design of the territorial scope of application, also future AI regulation would apply to “all relevant economic operators providing AI-enabled products or services in the EU, regardless of whether they are established in the EU or not.”[123] Conformity assessments would become mandatory for all economic operators of high risk AI applications regardless of their place of establishment. Mutual recognition agreements with third countries are foreseen as a way to selectively recognize conformity assessment conducted by third-country bodies. Moreover, the Commission signals that should an AI system not pass the conformity assessment it may need to be re-trained in the EU to ensure that for example the requirement on representativeness of data sets are being met.[124]

3. COMPARISON BETWEEN THE OPINION AND THE WHITE PAPER

The White Paper has been the subject of a public consultation which generated a very large number of submissions from a variety of stakeholders.[125] The Commission’s future direction for rule-making will predetermine how AI governance in the EU will look like. There are a few observations when comparing the recommendations of the Data Ethics Commission and the White Paper on Artificial Intelligence of the Commission.

First, the Commission’s White Paper has been criticized for not being ambitious enough.[126] Defining high-risk AI applications based on the cumulative criteria of predefined high-risk sectors and a high-risk AI application may be inflexible and cast the scope of application too narrow. Below the high-risk threshold many AI applications would not be covered by a future EU instrument. The White Paper does not seem to recognize risks posed by AI applications for groups and society at large, moreover it fails to gauge that many small risks can add up in widely used AI applications.

It follows as a consequence that, second, mandatory requirements foreseen in the White Paper for high-risk AI systems as regards information duties and data and record keeping duties would apply highly selectively. The criticality pyramid proposed in the Opinion by Germany’s Data Ethics Commission by contrast offers a more graduated approach to regulation than the proposal of the White Paper to focus exclusively on high-risk AI systems. The threshold for regulatory intervention recommended in the Opinion is significantly lower since not only high-risk applications would be addressed but already applications with some potential of harm have to comply with transparency and accountability

[121] European Commission, ‘White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)’ (n 2).

[122] ibid 20.

[123] ibid 22.

[124] ibid 23.

[125] Over 1250 replies were received, see European Commission (2020). Summary Report on the open public consultation on the White Paper on Artificial Intelligence. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68462

[126] See e.g. Federation of German Consumer Organisations (vzbv), ‘White Paper on Artificial Intelligence: Proposals of the Federation of German Consumer Organisations - Vzbv’ (2020).

standards. The establishment of a voluntary labelling scheme for no-high risk AI applications proposed by the Commission can hardly compensate for the lack of mandatory requirement at a medium risk category.

Also in substance, the White Paper is premised on imposing less stringent transparency requirements on economic operators. Beyond information duties for affected individuals, the White Paper does not foresee the publication of a risk assessment or conditional rights of access for journalistic and research purposes that have been recommended by the Data Ethics Commission. This would result in less public disclosure about the training and input data and the AI system's logic, its robustness, accuracy and fairness relative to a wider set of values. The ex ante conformity assessments favoured by the Commission would certainly involve an independent check of a given high-risk AI application but would lead to rather limited public documentation of AI systems risk assessment and management.

Also with regards to supervision and enforcement the recommendations by the Data Ethics Commission are more specific about technical requirements underscoring the need for standardized interfaces to live monitor high-risk AI systems or the retention of audit logs both for inspection by competent supervisory authorities.

The Commission's proposal in the White Paper however takes a clear stance on regulating economic operators in third countries providing AI-enabled products or services in the EU who would have to abide by EU rules on AI. This clarity must be welcomed.

THE EUROPEAN COMMISSION'S APPROACH TO HOLDING AI SYSTEMS ACCOUNTABLE SHOULD BE BOLDER

The White Paper on AI envisions selected information and documentation duties, ex ante conformity assessment and ex post supervision of high-risk AI systems. This approach disregards much needed transparency of ADM systems affecting individuals at moderate risk levels, the publication of ex ante impact assessments, enabling qualified transparency with standardized interfaces to carry out input/output audits and harnessing public interest research in justified cases.

III. AI RISKS ANTICIPATED FOR CONSUMER RIGHTS

Unless the expected EU legislation on AI governance declares consumer-facing AI as high-risk AI applications, European consumers have to contend themselves with existing consumer rights in the Union. This Section will provide an overview over the challenges for European consumer rights that are anticipated in the context of AI-powered consumer transactions, connected products and digital services from outside the EU. Anticipated challenges in enforcing consumer protections relate to the inscrutability of AI-related technologies and the cross-border supply of AI from businesses outside the EU.

The Section will examine European consumer rights other than individuals' privacy and data protection rights, which has been considered elsewhere.[127] Though European anti-discrimination law is not, strictly speaking, consumer protection law, a number of statutes require equality of treatment in contract law, thereby governing consumer transactions. This Section will sketch out two sets of challenges for European consumer rights: risks associated with AI in consumer products and services, and consumer risks in face of global electronic commerce.[128]

1. ANTI-DISCRIMINATION LAW

There is a body of anecdotal evidence and research underpinning that algorithmic bias poses a key challenge for our societies, for example several reported instances of differential treatment based on gender and race in the context of job advertisements and recruitment tools.[129] Already the training data can be a source of algorithmic bias when the data is not inclusive or representative for the population. Owing to the probabilistic methods deployed, algorithms are also quite prone to (re)produce some forms of statistical discrimination.[130] An AI system can even become biased by picking up on existing inequalities from previous practice:

> "The resulting predictions and recommendations extrapolate the past into the future, whereby existing social injustices can be obscured through incorporation into seemingly neutral technology, and potentially amplified."[131]

The following Section outlines which forms of discrimination are prohibited by EU law and what are the challenges for enforcing EU anti-discrimination laws in the context of AI.

a. Prohibited forms of discrimination

There is a defined range of characteristics on the basis of which discrimination is prohibited inside EU law. The most comprehensive anti-discrimination laws of the EU concern

---

[127] See for an overview Sartor (n 21); Svetlana Yakovleva and Kristina Irion, 'Pitching Trade against Privacy: Reconciling EU Governance of Personal Data Flows with External Trade' (2020) 0 International Data Privacy Law 1 <https://academic.oup.com/idpl/advance-article/doi/10.1093/idpl/ipaa003/5813832>.

[128] This Section benefits from the valuable research assistance of Anne van der Sangen, student of the masters programme of information law at Amsterdam Law School.

[129] See for examples Orwat (n 59) 30f.; Philipp Hacker, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law' (2018) 55 Common Market Law Review 1143; Frederik Zuiderveen Borgesius, 'Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence' [2020] International Journal of Human Rights 1 <https://doi.org/10.1080/13642987.2020.1743976>; Sandra Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' (2019) 35 Berkeley Technology Law Journal; Agnieszka Jabłonowska and others, 'Consumer Law and Artificial Intelligence' (2018); Orwat (n 59) 30f.

[130] See for a details Carsten Orwat, 'Risks of Discrimination through the Use of Algorithms' (2020) 25f. <https://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/EN/publikationen/Studie_en_Diskriminierungsrisiken_durch_Verwendung_von_Algorithmen.pdf?__blob=publicationFile&v=2>.

[131] Data Ethics Commission (n 3) 167.

nationality and place of residence, racial and ethnic origin, and gender.[132] EU anti-discrimination rules on nationality and place of residence are driven by internal market objectives. One example is the Regulation (EU) 2018/302 by which unjustified geo-blocking and other forms of discrimination based on the customers' nationality or place of residence are prohibited.[133]

Otherwise, the principle of non-discrimination is less coherently applied in the field of consumer law.[134] Anti-discrimination measures in consumer markets are most pronounced in sectors that are of social importance, such as labour, payment, insurance or banking, energy and electronic communications services as well as services of general economic interest, such as social protection, health care, and education.[135] Discrimination based on economic status is generally not covered which is to some extent compensated by consumer protection law's concern for protection of the economically weaker party of a transaction.

The literature anticipates a number of shortcomings with current EU rules on equal treatment in the face of algorithmic decision making and AI systems. Increasing algorithmic personalisation and marketing products and services to individual consumers could for instance escape those laws that are based on the assumption that consumer offerings must be available to the public at large.[136]

> "If there is individual communication with the customer, it cannot be regarded as 'general conditions of access made available to the public at large'. Individual communication does not mean that the conditions are available "to the public at large"."[137]

Besides, current anti-discrimination law does not fully apply to free service offering that are commonplace online where revenue is made from different markets, such as online advertisements. Anti-discrimination statutes still require a monetary counterperformance for finding that a service contract was concluded between the provider of a digital service and the user.[138] This leads to a gap of protection in the context of free online service which has already been closed with most other EU consumer protection law,[139]

---

[132] Equal treatment based on nationality or residence: European Parliament and the Council, Directive 2006/123/EC of 12 December 2006 on services in the internal market, ELI: http://data.europa.eu/eli/dir/2006/123/oj; European Parliament and the Council, Regulation (EU) 2018/302 of 28 February 2018 on addressing unjustified geo-blocking and other forms of discrimination based on customers' nationality, place of residence or place of establishment within the internal market, ELI: http://data.europa.eu/eli/reg/2018/302/oj;

Equal treatment based on racial and ethnic origin: Council, Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, ELI: http://data.europa.eu/eli/dir/2000/43/oj;

Equal treatment based on gender: Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation, ELI: http://data.europa.eu/eli/dir/2006/54/oj.

[133] Regulation (EU) 2018/302 ibid. (n 127).

[134] Jabłonowska and others (n 129) 21.

[135] See for an overview Jabłonowska and others (n 129).

[136] Hacker (n 129) 1154f.

[137] Hans Schulte-Nölke and others, 'Discrimination of Consumers in the Digital Single Market' (2013) 45 <http://www.europarl.europa.eu/RegData/etudes/STUD/2013/507456/IPOL-IMCO_ET(2013)507456_EN.pdf>.

[138] See Hacker (n 129) 1155.

[139] See European Parliament and of the Council, Directive (EU) 2019/770 of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services, ELI: http://data.europa.eu/eli/dir/2019/770/oj, Article 3 (1); and as part of the "New Deal for Consumers" initiative, see European Parliament and of the Council, Directive (EU) 2019/2161 of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules, ELI: http://data.europa.eu/eli/dir/2019/2161/oj.

Researchers point out that EU and member states approach to non-discrimination may be too rigid to prevent differential treatment through AI, in cases where it is based on inferences that do not correlate with protected grounds.[140] For example, an algorithm may identify a peculiar mix of traits and cues which narrows in on individuals who share characteristics that are outside the scope of non-discrimination statutes.[141]

Whereas direct discrimination based on one of the qualified characteristics, such as ethnic origin or religious belief, can only be justified based on very limited grounds, justifying indirect discrimination, such as ability to pay or social status, is more problematic. Several anti-discrimination directives provide that indirect discrimination can objectively be justified when the practice pursues a legitimate aim, and the means of achieving that aim are appropriate and necessary. Economic actors could invoke recognized levels of algorithmic accuracy to justify indirect discrimination as objective and legitimate.[142]

b. Price discrimination

Price discrimination relies on sellers sorting buyers in terms of willingness to pay; consumers' that are willing to pay more are charged a higher price.[143] Price discrimination is nothing new. However, AI supercharges the ability of digital retailers to engage in much more precise, targeted and dynamic forms of price discrimination. Through tracking of online behaviour, providers can estimate consumers' potential willingness to pay and even recognise emotion.[144]

The GDPR regulates the legitimacy and conditions of handling consumers' personal data which are the input for personalizing pricing techniques.[145] However, the output, i.e. personalised prices, is perfectly acceptable from the perspective of EU consumer protection law, unless it qualifies as a prohibited form of discrimination as outlined above.[146] Both, the Unfair Terms Directive and the Unfair Commercial Practices Directive leaves traders free to set prices.[147] The 2019 "New Deal for Consumers" initiative at least introduced, as a new information requirement, that traders have to inform the consumer when the price is personalised on the basis of automated decision-making.[148] Only the Service Directive includes a general prohibition on price discrimination based on nationality and place of residence as regards services which are made available to the public at large.[149]

---

[140] Zuiderveen Borgesius (n 129) 15.

[141] Wachter (n 129).

[142] Hacker (n 129).

[143] Frederik Zuiderveen Borgesius and Joost Poort, 'Online Price Discrimination and EU Data Privacy Law' (2017) 40 Journal of Consumer Policy 347.

[144] Christopher Townley, Eric Morrison and Karen Yeung, 'Big Data and Personalised Price Discrimination in EU Competition Law' (2017) 36 Yearbook of European Law.

[145] Note that Article 22 of the GDPR covers automated individual decision-making and profiling, however, the scope of this provision is limited to "a decision based solely on automated processing, including profiling, which produces legal effects …". It will not suffice to deal with ADM properly, see e.g. Data Ethics Commission (n 3) 185; Martini (n 53) 10f.

[146] Agnieszka Jabłonowska and others, 'Consumer Law and Artificial Intelligence Challenges to the EU Consumer Law and Policy Stemming from the Business' Use of Artificial Intelligence' (2018) EUI Working Paper 50f.

[147] Article 4 (2) of the Unfair Terms Directive excludes the adequacy of the price and remuneration from the assessment of the unfair nature of contract terms as long as they are in plain intelligible language. Council, Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts (unofficial consolidated version), ELI: http://data.europa.eu/eli/dir/1993/13/2011-12-12.

As long as traders refrain from misleading consumers over the price or the manner in which the price is calculated, or the existence of a specific price advantage traders can charge different prices, see Article 6 (1) of the Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market ('Unfair Commercial Practices Directive'), ELI: http://data.europa.eu/eli/dir/2005/29/oj.

[148] Article 4 (a) (ii) of Directive (EU) 2019/2161.

[149] Article 20 (2) of Directive 2006/123/EC. However, this Directive does not preclude the possibility of providing for differences in the conditions of access where those difference are directly justified by objective criteria.

Irrespectively, consumers are sceptical of personalized pricing which is regarded unfair and even manipulative.[150] Moreover, personalised pricing could be problematic "if certain groups in society structurally pay more for goods and services."[151] It can lead to the economic exploitation of vulnerable individuals, i.e. "those on low incomes, the elderly, the digitally disempowered (including those with no ready access to the Internet), and the poorly educated."[152]

> "Because personalization strategies serve to isolate individual consumers from each other, they thereby erode consumers' power to act collectively in ways that might serve their interests as a whole."[153]

For the time being, the socio-economic effects of personalised prices and their impact on social solidarity and cohesion should be kept under scrutiny.[154] Keeping tabs on personalised pricing practices would however require real-world data about such practices and its analysis by authorities and consumer protection organisations.

c. Enforcement and burden of proof

AI's characteristic opacity (or 'black-box-effect') and ADM's discretionary application makes it harder to detect and ascertain unlawful practices. While EU legislation remains applicable irrespective of the involvement of AI, it is important to assess whether it can be adequately enforced.[155] Some prior knowledge and evidence about how an AI system works and makes decisions will be key to the enforcement of non-discrimination rights through regulators and in the courts.

When taking legal action against a discriminatory practice involving ADM, consumers or their representative organisations would find it hard to satisfy the burden of proof required by the legislation.

> "From the perspective of the persons concerned, the poor traceability of the effects of algorithms makes it difficult or even impossible for the persons concerned to demonstrate that they have suffered discrimination due to algorithms."[156]

Consider for instance that claiming indirect discrimination has to demonstrate "that a seemingly neutral rule, practice, or decision disproportionately affects a protected class and is thus prima facie discriminatory."[157]

Gathering prima facie evidence would require information about the treatment of other individuals which is not easy to obtain from public sources. Successful explorations of algorithmic discrimination so far are based on extensive empirical studies and algorithm audits, which are often not within the reach of common users.[158] In face of the considerable uncertainties, affected individuals could be less likely to initiate civil law litigation against suspected unequal treatment in relation to consumer transactions and shoulder

---

[150] Zuiderveen Borgesius and Poort (n 33) 349; European Commission, Staff Working Document, Sec. 5.2.12.

[151] Zuiderveen Borgesius (n 129) 15; Sartor (n 21) 36.

[152] Karen Yeung, 'Five Fears about Mass Predictive Personalization in an Age of Surveillance Capitalism' (2018) 8 International Data Privacy Law 258, 262.

[153] ibid 261.

[154] Yeung (n 39) 34f.

[155] European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)' (n 2).

[156] Orwat (n 59) 72.

[157] Frederik J Zuiderveen Borgesius, 'Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence' (2020) International Journal of Human Rights, 6 <https://doi.org/10.1080/13642987.2020.1743976>; Wachter (n 32).

[158] Orwat (n 59) 72; Sandvig and others (n 54); Bodo and others (n 54).

the associated risk of legal costs. Experts propose lowering the burden of proof for individual redress against unequal treatment:

> "… legislators should enact legislation clarifying the requirements for providing proof of discrimination by operators of algorithmic systems and lower such requirements further for affected parties as needed."[159]

European anti-discrimination statutes differ in their approach to public enforcement, collective action and redress mechanisms.[160] In any case EU law provides for designated anti-discrimination bodies in the Member States to promote the equal treatment of all persons without discrimination and to analyse, observe and provide support.[161] Even where these equality bodies have no investigatory powers, e.g. in Germany, they can carry out research on discrimination in ADM. Yet, there is currently too little experience or practice to engage at the level of AI technology or to conduct external algorithmic audits.

2. CONSUMER PROTECTION LAW

Ensuring a high level of consumer protection is an important EU policy objective.[162] EU consumer protection law seeks to eliminate barriers to the internal market by assisting consumers and protecting them from various risks in consumer transactions that are hard to cope with as the weaker party.[163] The body of EU consumer protection laws encompasses a large number of instruments dealing with specific consumer rights, such as the provision of pre-contractual information and a right of withdrawal, unfair terms and commercial practices, misleading and comparative advertising, product safety and liability, among others.[164] While consumers expect the same level of safety and respect of their rights whether or not a product or a system relies on AI, it can make the application and enforcement of this legislation more difficult. The following sub-sections will provide an overview of the issues that AI raises for consumer protection, namely undue influence and manipulation, product safety and liability.[165]

a. Undue influence and manipulation

AI can be used to influence consumers with the aim of steering them towards making choices that may not serve their best interests.[166] This form of influence utilises insights from behavioural sciences to identify which emotions make consumers buy certain products.[167] AI is designed to go deeper into the needs and interest of individual costumers. For some consumers this might be beneficial, but there is the risk of undue influence and even manipulation:

---

[159] Data Ethics Commission (n 3) 195.

[160] Orwat (n 59) 84.

[161] In Article 13 of Directive 2000/43/EC, Article 12 of Directive 2004/113/EC, and Article 20 of Directive 2006/54/EC.

[162] Charter of Fundamental Rights of the European Union (Charter) Article 38.

[163] The EU shall specifically contribute to "protecting the health, safety and economic interests of consumers, as well as to promoting their right to information, education and to organise themselves in order to safeguard their interests." Treaty on the Functioning of European Union (TFEU) Article 169 (1).

[164] Most EU legislation in the field of consumer protection takes the form of directives which need to be transposed into Member State's national law. As a shared competence of the Member States and the EU, Member States can maintain or introduce more stringent protective measures than provided for by the EU harmonization directive (TFEU Article 169 (4)). See for more background Jana Valant, 'Consumer Protection in the EU: Policy Overview' [2015] European Parliamentary Research Service <https://www.europarl.europa.eu/thinktank/en/document.html?reference =EPRS_IDA(2015)565904>; Stephen Weatherill, *EU Consumer Law and Policy* (Edward Elgar Publishing 2005).

[165] BEUC, 'Survey: Consumers See Potential of Artificial Intelligence but Raise Serious Concerns' *Press Release* (Brussels, September 2020).

[166] Sartor (n 21) 3; Daniel Susser, Beate Roessler and Helen Nissenbaum, 'Online Manipulation: Hidden Influences in a Digital World' (2019) 4 Georgetown Law Technology Review 1.

[167] Maurits Kaptein, *Persuasion Profiling: How the Internet Knows What Makes You Tick* (Amsterdam 2015).

> "Consumers are in a weak position when facing automated persuaders, which have access to a huge amount of knowledge, can effortlessly deploy unlimited computational power, and can frame the choice and information environment that is available to consumers."[168]

It can be questioned whether the Unfair Commercial Practices Directive is up to the challenge to guarantee consumers' freedom of choice and conduct in an AI-driven environment.[169] According to the Articles 8 and 9 of this Directive marketing must not involve aggressive commercial practices or exert undue influence on consumers. The European Commission provides as an example putting a consumer under time pressure when buying a flight ticket by falsely claiming that only a few tickets are left available.[170] Such deceitful marketing could be deemed an unfair commercial practice in the meaning of the Directive.[171] This example demonstrates that in order to show that there has been an unfair commercial practice additional information is necessary, which typically only the trader has.

ADM-powered consumer markets are likely to produce information asymmetries that make consumers the significantly weaker party before and during a transaction.

> "Overall, the extent to which the weaker party protection is revitalized through the use of big data and learning algorithms by the businesses appears as one of the most pertinent questions to be addressed in the course of further research on consumer law and AI."[172]

Simply supplying consumers with additional information has known limitations since it takes time and effort to read and understand the information as well as to act upon it. Making consumers bear the burden of managing their exposure to ADM systems is in itself not effective as a protection for the weaker party. From the field of personal data protection it is already well known that individuals are overwhelmed by too many and too long privacy notices and tend to disregard them, so expecting them to manually opt-out from an infinite number of purposes is not realistic.[173]

b. Product safety and liability

Product safety and product liability legislation are two complementary mechanisms that ensure high levels of safety of products marketed in the Union. EU product safety legislation has to ensure that products placed on the market are safe and that such products can circulate freely in the internal market. Its centrepiece, the General Product Safety Directive, covers any product intended for consumers use or likely to be used by consumers, including in the context of providing a service. The Directive requires that every product placed on the market complies with relevant national product safety standards of the Member State in whose territory the product is marketed.[174]

---

[168] Sartor (n 43) 5.

[169] European Parliament and the Council, Directive 2005/29/EC (n 35).

[170] European Commission, Staff Working Document Guidance on the Implementation/Application of Directive 2005/29/EC on unfair commercial practices, SWD (2016) 163 final, Section 5.2.13 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016SC0163>.

[171] European Parliament and the Council, Directive 2005/29/EC (n 35), Article 6(1)(a) and Annex I No 7.

[172] Jabłonowska and others (n 146) 15.

[173] Daniel J Solove, 'Privacy Self-Management and the Consent Dilemma' (2013) 126 Harvard Law Review 1880; Manon Oostveen and Kristina Irion, 'The Golden Age of Personal Data: How to Regulate an Enabling Fundamental Right?' in Mor Bakhoum and others (eds), *Personal Data in Competition, Consumer Protection and IP Law - Towards a Holistic Approach?* (Springer Press 2018).

[174] European Parliament and the Council, Directive 2001/95/EC of 3 December 2001 on general product safety (unofficial consolidated version) ELI: http://data.europa.eu/eli/dir/2001/95/2010-01-01.

Emerging technologies, such as AI, connected devices and the Internet of Things (IoT) are transforming the characteristics of many products in the market, thus bringing new challenges and risks related to product safety. In its report on the safety and liability aspects of AI the Commission summarizes the gaps in the application of the current legislative framework as follows:

> "While the Union product safety legislation takes into account the safety risks stemming from software integrated in a product at the time of its placing on the market and, potentially subsequent updates foreseen by the manufacturer, specific and/or explicit requirements on standalone software could be needed."[175]

Next to the need to update legislation for standalone digital services, the report considers a range of challenges for the product safety framework, notably connectivity, autonomy, opacity, complexity and responsibility in complex value chains. While the overall safety concept encompasses protection against all kinds of risks arising from the product, certain safety aspects would require an update of the legislation. For example, that a risk assessment must be repeated for self-learning products because they can affect product safety over time.[176] Overall, the report makes a case for introducing AI regulation to ensure human oversight and introduce requirements for transparency of algorithms, as well as for their robustness, accountability and unbiased outcomes.[177]

Under the Product Liability Directive the producer is liable for damage caused by a defect in his product.[178] As the General Product Safety Directive, the Product Liability Directive also covers products but not standalone services, which affects consumers' ability to recover damages incurred as a result of faulty digital services. Only when a damage is due to a defective product, used in the provision of a service, it will be recoverable under the Product Liability Directive. When software is supplied over the Internet, however, potential defects do not fall within the scope of this Directive.

It is possible that AI-driven technology fails, either unintentionally or by design, leading to property damages and/or economic loss. The report on the safety and liability aspects of AI flags that certain characteristics of AI can reduce the effectiveness of EU and national liability frameworks:

> "Some of these characteristics could make it hard to trace the damage back to a human behaviour, which could give grounds for a fault-based claim in accordance with national rules. This means that liability claims based on national tort laws may be difficult or overly costly to prove and consequently victims may not be adequately compensated."[179]

Additional legal uncertainty can arise over how EU and national liability laws would be applied to damages caused by AI.[180] Operationalising legal concepts, such as harm and fault or establishing causality between the fault of the liable person and damage can be difficult, especially in situations in which an AI is involved. Satisfying the burden of proof

---

[175] European Commission, Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64 final <https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf>.

[176] Ibid. 7.

[177] Ibid. 9.

[178] Council, Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products (unofficial consolidated version), ELI: http://data.europa.eu/eli/dir/1985/374/1999-06-04.

[179] European Commission, Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics (n 46) 13.

[180] Ibid., 12f.

when claiming compensation for damages caused by an AI application, self-leaning or not, can be difficult for consumers:

> "The need to understand the algorithm and the data used by the AI requires analytical capacity and technical expertise that victims could find prohibitively costly. In addition, access to the algorithm and the data could be impossible without the cooperation of the potentially liable party."[181]

It may be necessary, following the Commission, to alleviate or even reverse the burden of proof under national liability law which currently places the responsibility to submit evidence and establish causation on the claimant.[182]

c. Private and public enforcement

Consumer rights are asserted via private and public enforcement. A consumer can for example initiate civil law litigation at a domestic court to seek redress for an infringement of individual consumer rights. One concern is that the information asymmetries between the operator and the user may deter individuals to seek access to justice. Another concern is that the corrective effect of individual redress mechanisms would be almost imperceptible relative to a large scale ADM system. The reason is that civil law remedies are only binding in the relationship between the parties to the proceeding. Thus, obtaining a court ruling against a certain unlawful business practice involving an AI system would be the proverbial drop on a hot stone.

Only collective redress and class actions can potentially help overturn an AI application that violates individual rights at scale. Representation of the collective interests of consumers can be powerful means in private enforcement of consumer protection law. The EU has recently adopted a new Directive that strengthens representative actions for the protection of the collective interests of consumers which Member States have yet to be implement in their national laws by the end of 2022.[183] According to this Directive qualified consumer protection bodies will have legal standing in domestic courts for representative action that can be for redress or injunction measures. Private enforcement by means of representative action would still need to provide prima facie evidence that a specific ADM system infringes consumer rights.

Although national consumer protection authorities are equipped with investigatory powers, they face an uphill battle in supervising increasingly ubiquitous AI in consumer markets. Contextual information that would be necessary to assess whether a commercial practice is unfair is often not in plain sight, but can require meticulous investigations. Typically, these authorities' oversight and enforcement activities do not scale but are conducted case by case. Even if they step up their capabilities to assess and audit ADM systems it may not be feasible to keep abreast with the diffusion of AI in consumer markets and mass-personalisation of consumer advertisement, transactions, and recommendations, as well as monitor the effects of personalised pricing practices.

3. CONSUMER RIGHTS IN GLOBAL ELECTRONIC COMMERCE

Another set of challenges, however, emanate from global electronic commerce connecting EU consumers with traders from all over the world. These challenges are not specifically attributable to AI but they can arise as a consequence of buying cross-border from online retailers and using services from providers outside the EU. EU consumer rights in

---

[181] Ibid. 15.

[182] Ibid., 14; see also Expert Group on Liability and New Technologies – New Technologies Formation, *Liability for Artificial Intelligence Report from the Expert Group on Liability and New Technologies-New Technologies Formation* (European Union 2019) 30.

[183] European Parliament and the Council, Directive (EU) 2020/1828 of the European Parliament and of the Council of 25 November 2020 on representative actions for the protection of the collective interests of consumers and repealing Directive 2009/22/EC, ELI: http://data.europa.eu/eli/dir/2020/1828/oj.

the form of information duties, protection against unfair commercial practices, rights of withdrawal and redress can be very difficult to obtain with a trader outside the EU.[184] The vastness of the global electronic market place is not paired with effective consumer protection mechanisms.

Research on consumer protection in global electronic commerce suggests that EU consumers are hesitant of making purchasing from online retailers outside the EU, though their numbers are growing in the past years, mostly due to the rise of online market places.[185] The lack of trust is mostly attributed to a lack of confidence in cross-border electronic commerce on part of the consumers. While this holds true for consumer goods, EU consumers very frequently use digital services supplied from economic operators outside the EU, for instance digital apps, virtual personal assistance and digital content services.[186] Here consumers do not display similar reservations because these digital services satisfy consumer demand and can be supplied free of charge.

a. European private international law

European private international law is subscribed to the principle of the protection of a weaker party. Following EU regulation determining jurisdiction, consumer contracts, subject to certain qualifying conditions, are "governed by the law of the country where the consumer has his habitual residence."[187] Consumer contracts are moreover protected against an adverse choice of law that would deprive the consumer of the protection afforded by law.[188] EU consumer protection law knows other consumer conflict-of-law rules, for example the Unfair Terms Directive provides:

> "Member States shall take the necessary measures to ensure that the consumer does not lose the protection granted by this Directive by virtue of the choice of the law of a non-Member country as the law applicable to the contract if the [consumer] has a close connection with the territory of the Member States."[189]

Consumer conflict-of-law rules have helped to default consumer protection litigation to the Member State of the consumer inside the Union.[190] They have potential to protect consumers against the enforcement of foreign judgements which are in conflict with EU consumer conflict-of-law rules. However, they can hardly commit commercial actors which operate out of third countries to adhere to EU consumer contract law in the first place. While some economic operators with a strong foothold in EU markets can have an incentive to incorporate EU consumer protection law upfront, there are large swaths of operators who will not.[191]

---

[184] Julie Hunter and others, 'The Challenge of Protecting EU Consumers in Global Online Markets' (2017) 5 <http://www.beuc.eu/publications/beuc-x-2017-122_the_challenge_of_protecting_eu_consumers_in_global_online_markets.pdf>.

[185] ibid 13.

[186] See e.g. Forbrukerrådet, 'Appfail: Threats to Consumers in Mobile Apps' (2016) <https://fil.forbrukerradet.no/wp-content/uploads/2016/03/Appfail-Report-2016.pdf>; European Commission, 'The Rise of Virtual Personal Assistants' (2018) <https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/Virtual personal assistants_v1.pdf>.

[187] European Parliament and the Council, Regulation (EC) No 593/2008 of 17 June 2008 on the law applicable to contractual obligations (Rome I), OJ 4.7.2008, L177/6, Article 6 (1). See related Regulation (EU) No 1215/2012 of the European Parliament and of the Council of 12 December 2012 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters (recast) (unofficial consolidated version), ELI: http://data.europa.eu/eli/reg/2012/1215/2015-02-26, Article 17.

[188] Ibid. Article 6 (2).

[189] Council, Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts (n 35), Article 6 (2).

[190] E.g. CJEU, Judgment of 28 July 2016, case C-191/15 (*Verein für Konsumenteninformation*), EU:C:2016:612.

[191] See e.g. for apps and cloud services Norwegian Consumer Council, 'Hazy Terms in the Cloud' (2014) 14 <http://www.forbrukerradet.no/annet/tester-og-kj%25C3%25B8petips/unders%25C3%25B8kelser/hazy-terms-in-the-cloud>; Forbrukerrådet, 'Appfail: Threats to Consumers in Mobile Apps' (2016).

b. Cross-border enforcement cooperation

International guidance of the OECD and the United Nations enshrine the equality of consumer protection as a key principle.[192] Consumer rights and protection however remain a fragmentated landscape since countries and the EU differ in their approaches, priorities and legal traditions.[193] There are efforts to step up consumer protection enforcement cooperation and cross-national assistance of national consumer protection authorities. The 2003 OECD Guidelines for Protecting Consumers from Fraudulent and Deceptive Commercial Practices across Borders are a case in point.[194] The International Consumer Protection and Enforcement Network (ICPEN) is a network 65 consumer protection authorities which mainly engages in knowledge exchange and best practice sharing.[195]

A recent OECD report finds that the majority of the surveyed countries have the legal authority to provide or facilitate remedies for foreign consumers.[196] However, even where enabling legislation exists, the report continues, considerable barriers to cross-border co-operation remain:

> "The most important factor is a lack of adequate resources in consumer protection enforcement authorities. Around 70% of countries reported that inadequate resources are always (18%) or frequently (50%) barriers to cross-border co-operation."[197]

The upshot is that cross-border enforcement cooperation in consumer protection is still in its infancy. Protecting consumers in the global digital marketplace would require exponentially increased enforcement co-operation in order to be effective. Third countries' consumer rights in connection with AI technology are not yet fully formed either and provide thus an unlikely fall-back for EU consumers.

4. ENFORCING CONSUMER RIGHTS AGAINST HARMFUL AI

Digital technologies have become inseparable from our daily experience as consumers which can bring new challenges for consumer rights.[198] Consumer facing AI markets display sizeable information asymmetries between the controllers of AI technology, on the one side, and, on the other side, individual consumers. It is important to distinguish two sets of challenges, namely anticipated risks of AI technology for European consumer rights and those risks stemming from global electronic commerce.

Under the first set of challenges, consumer rights can be affected by ADM in the context of consumer marketing, transactions, products and services. The risks for consumer rights concern both questions of substantive law and procedural aspects but also the limited enforcement capacity of anti-discrimination and consumer protection authorities in the member states. A recurring theme that negatively affects all consumer rights is the difficulty of satisfying the burden of proof that an AI system is faulty, biased or unfair. This

---

[192] OECD (2016), Consumer Protection in E-commerce: OECD Recommendation (Paris: OECD Publishing) <http://dx.doi.org/10.1787/9789264255258-en>; United Nations (2015), United Nations Guidelines for Consumer Protection, adopted by the General Assembly in resolution 70/186 of 22 December 2015 (New York and Geneva: United Nations) <http://unctad.org/en/PublicationsLibrary/ditccplpmisc2016d1_en.pdf>.

[193] Monique Goyens, 'Effective Consumer Protection Frameworks in a Global and Digital World' (2020) 43 Journal of Consumer Policy 195, 196 <http://link.springer.com/10.1007/s10603-019-09423-2>.

[194] OECD, Recommendation of the Council concerning Guidelines for Protecting Consumers from Fraudulent and Deceptive Commercial Practices across Borders <https://legalinstruments.oecd.org/public/doc/184/184.en.pdf>.

[195] See ICPEN website at <https://icpen.org/>.

[196] OECD (2018), "Consumer protection enforcement in a global digital marketplace", OECD Digital Economy Papers, No. 266 (Paris: OECD Publishing) <http://dx.doi.org/10.1787/f041eead-en>.

[197] Ibid. 6.

[198] Marco Lippi and others, 'The Force Awakens: Artificial Intelligence for Consumer Law' (2020) 67 Journal of Artificial Intelligence Research 169.

can be resolved by either reversing the burden of proof or escalating algorithmic transparency for users of ADM systems hand-in-hand with meaningful oversight and verification mechanisms.

Under the second set of challenges, there are risks for consumer rights arising from global electronic commerce when economic operators in third countries are not bound by or do not comply with EU law. The means to enforce EU consumer rights vis-à-vis economic operators outside the EU are burdensome and outcomes highly uncertain. Existing mechanisms for cross-border cooperation outside the EU are inadequate in scope and scale to ensure effective enforcement of cross-border consumer rights cases. Before AI even enters the picture there is a mismatch "between the twentieth-century consumer law and twenty-first-century market developments, qualified by globalization and digitalization."[199]

These cross-border challenges multiply when AI's characteristic opacity (or 'black-box-effect')[200] can obstruct infringement detection, lawfulness and enforcement of EU consumer protection standards in global electronic commerce. The questions that are bound to arise in cross-border consumer protection environments are thorny and unresolved: how can the cooperation of a non-EU business that has no branch, agency or other establishment in the EU be ensured? What if such cooperation would be necessary to establish a claim of algorithmic bias or manipulation contrary to European law, or that a damage has occurred as a consequence of a faulty AI system? What if consumer protection authorities in the member states cannot compel a non-EU operator to provide information that is critical for enforcement?

CONSUMER PROTECTION NEEDS ENFORCEMENT THAT LIVES UP TO THE RISKS OF ADM AND GLOBAL ELECTRONIC COMMERCE

Public and private enforcement face significant difficulties in overcoming AI's characteristic opacity in order to enforce EU consumer rights. If future EU rules on AI governance will not apply to consumer-facing AI, these difficulties will not be mitigated. Besides, our current system of enforcement in individual cases after an infringement has occurred is not able to cope with digital consumer markets increasingly powered by AI systems and mass-personalisation. Proposals to resolve challenges posed by ADM systems include alleviating the burden of proof, enhancing regulatory enforcement capacity as well as leveraging collective redress and public scrutiny of AI systems. Additional safeguards will be important in situations where EU consumers are at the receiving end of ADM systems from businesses that operate from outside the EU.

[199] Goyens (n 193).

[200] After the seminal book by Pasquale (n 17).

IV. EU TRADE LAW OBLIGATIONS, AI AND A NEW SOURCE CODE DISCIPLINE

The EU is committed to promoting international trade in the context of the rule-based multilateral trading system. The Commission's 2015 strategy "Trade for All" was based on the premise that EU's trade and investment policy must further embrace today's economic system which is global and digital at its core.[201] By today the geopolitical environment has changed dramatically:

> "Tensions among the major global economies, a rise of unilateralism and economic nationalism, stronger involvement of the state in the economy, the weaponisation of trade policy for economic or geopolitical objectives–all these factors have led to a weakening of global governance structures generally, and the multilateral rules-based order in particular."[202]

A closer look at international economic policy relating to digital services reveals controversial but interconnected issues, such as reforming taxation rules for digital services[203] parallel to the ongoing WTO negotiations on trade-related aspects of electronic commerce.[204] China and recently the U.S. increasingly resort to unilateral restrictions of trade in data, digital services, technologies and AI on grounds of national security against one another.[205]

The EU is navigating a delicate balance as to "how EU trade policy can support the digital transition and help secure the EU's position in the digital sphere in the long-term."[206] This Section reviews the EU trade law obligations under the WTO General Agreement on Trade in Services (GATS) before taking a closer look at the EU Proposal for WTO Disciplines and Commitments Relating to Electronic Commerce tabled at the WTO for current negotiations.[207] The core of this Section analyses the scope and meaning of a new discipline on source code of software which is backed by the EU proposal.

1. EU'S COMMITMENTS UNDER THE GATS

A brief overview on EU's commitments under the GATS is appropriate, since the GATS provides an important reference framework for the WTO electronic commerce negotiations under way. Both the EU and its member states are founding members of the WTO and parties to the GATS. This is the first multilateral treaty on the liberalization of international trade in services which entered into force in 1995, as a result of the Uruguay

---

[201] European Commission, *Trade for All: Towards a More Responsible Trade and Investment Policy* (2015) 7 <http://ec.europa.eu/trade/policy/in-focus/new-trade-strategy/%5Cnhttp://trade.ec.europa.eu/doclib/docs/2015/october/tradoc_153846.pdf>.

[202] European Commission, 'A Renewed Trade Policy for a Stronger Europe: Consultation Note' (2020); Anon., 'It's the End of the World Trade Organisation as We Know It' *The Economist* (2019).

[203] OECD, "International community renews commitment to multilateral efforts to address tax challenges from digitalisation of the economy", 31 January 2020 https://www.oecd.org/tax/international-community-renews-commitment-to-multilateral-efforts-to-address-tax-challenges-from-digitalisation-of-the-economy.htm; Ryan Heath, "EU pushing ahead with digital tax despite U.S. resistance, top official says," Politico, 23 June 2020 <https://www.politico.com/news/2020/06/23/eu-digital-tax-united-states-336496>.

[204] WTO (n 23); Jane Kelsey and others, 'How " Digital Trade " Rules Would Impede Taxation of the Digitalised Economy in the Global South' (2020) <https://twn.my/title2/latestwto/general/News/Digital Tax.pdf>.

[205] See Michael R Pompeo, 'Announcing the Expansion of the Clean Network to Safeguard America's Assets' *Press Release* (Washington DC, 2020) <https://www.state.gov/announcing-the-expansion-of-the-clean-network-to-safeguard-americas-assets/>; Marietje Schaake, 'EU Risks Being Dethroned as World's Lead Digital Regulator' *Financial Times* (2020) <https://www.ft.com/content/233a7dde-30f1-45bb-acdb-926b6969b952>.

[206] European Commission, 'A Renewed Trade Policy for a Stronger Europe: Consultation Note' (n 202).

[207] WTO (n 7).

Round negotiations.The GATS aims for the expansion of international trade in services through the elimination of trade barriers.

The preamble to the GATS recognises "the right of Members to regulate, and to introduce new regulations, on the supply of services within their territories in order to meet national policy objectives".[208]

a. A WTO member's autonomy to regulate

Even though the GATS does not have the deregulation of services as an objective,[209] a member's autonomy to regulate is not without boundaries. Some of these boundaries relate to the rule that a member's domestic regulation affecting trade in services must be consistent with the GATS and applied in a non-discriminatory fashion.[210] The margin of manoeuvrability left to a WTO member is then further prescribed by its individual schedule of commitments in the disciplines of market access and national treatment (i.e. the principle of non-discrimination between national and foreign products or services). In practice, a WTO member's autonomy to adopt a GATS-inconsistent regulation is then confined to the limits of the GATS general exceptions (see below).

In how far a WTO member's autonomy to regulate is determined by its individual schedule of commitments can be illustrated based on the commitments the EU inscribed in relation to Computer and Related Services category. It is worth bearing in mind that at the time of the Uruguay Round of negotiations (from 1986 to 1993) the economic relevance of this sector was modest and regulatory intervention at the domestic level negligible. The EU inscribed far-reaching commitments for both market access and national treatment for all sub-sectors in the service category Computer and Related Services.[211]

> "The implications of these commitments are real and the wiggle-room available for domestic regulators is severely constrained."[212]

The GATS which protects cross-border trade in services incidentally also protects how service suppliers integrate data flows and processing operations into their ordinary course of business.[213] Domestic regulation that affects the conditions of supplying digital services can quickly turn into some kind of a behind-the-border barrier to trade. Regulatory interventions at the level of digital architecture, while not discriminating on the face of it, can modify the conditions of competition in favour of domestic services.[214] A domestic measure that is thus found to discriminate against foreign companies and their services violates a GATS member's commitment on national treatment. The justification of a GATS-inconsistent measure will be discussed below.

---

[208] GATS Preamble.

[209] Peter van den Bossche and Werner Zdouc, *The Law and Policy of the World Trade Organization* (3rd edition, Cambridge University Press 2014) 514.

[210] Pursuant to GATS Articles VI(1) and XIV.

[211] The EU has listed no limitations for the modes of supply 1, 2 and 3 (cross-border, consumption abroad and commercial presence) in the sub-sectors a) consultancy services related to the installation of computer hardware, b) software implementation services, c) data processing services, d) data base services, maintenance and repair, as well as other e) computer services, see WTO, European Communities and their Member States, Schedule of Specific Commitments, Trade in Services, GATS/SC/31 (1997), s. II. 1. B a) to e).

[212] Mira Burri, 'Current and Emerging Trends in Disruptive Technologies : Implications for the Present and Future of EU's Trade Policy' (European Parliament 2017) 16 <http://www.europarl.europa.eu/RegData/etudes/STUD/2017/603845/EXPO_STU(2017)603845_EN.pdf>.

[213] Irion and Williams (n 18) 21.

[214] See e.g. Daniel Crosby, 'Analysis of Data Localization Measures Under WTO Services Trade Rules and Commitments' (2016) <https://e15initiative.org/wp-content/uploads/2015/09/E15-Policy-Brief-Crosby-Final.pdf>.

b. AI trade within the scope of the GATS

Digital services that incorporate AI into their software architecture are presumed to be already covered by the GATS.[215] Even though there are enduring questions regarding the proper service classification and the interpretation to be given to a member's scheduled commitments in a digital context, WTO adjudicating bodies have consistently found digital commercial activities to be covered by the GATS. Adding AI to the cross-border supply of digital services would not make a difference inside the GATS, because its service categories are considered as a proxy for rule application.

As a rule-of-thumb a digital service with a clear-cut analogue legacy would be subsumed under a generic entry in the service classification list.[216] For example, ML used in real-time bidding in online advertising is classified as Advertising Service, and so triggering attendant commitments in a GATS member's individual schedules. A digital service without a fitting analogue legacy is more likely to be covered under one of the Computer and Related Services sub-sectors. For example, an online search engine is presumably classified under the Data Processing Services sub-sector.[217]

The conclusion that the GATS presumptively governs AI should not come as a surprise to trade law experts. What is more astonishing is the lack of a broader understanding and public discourse about trade law's genuine role in the facilitation of cross-border trade in AI and the proper impact of the WTO's electronic commerce negotiations discussed below.

2. EU PROPOSAL FOR A WTO AGREEMENT ON ELECTRONIC COMMERCE

In 1998, WTO members agreed to launch a Global Work Programme on Electronic Commerce that would examine all trade-related issues relating to cross-border electronic commerce.[218] The work programme of the same year defines electronic commerce as "the production, distribution, marketing, sale or delivery of goods and services by electronic means."[219] With the exception of periodically prolonging the moratorium on payment of custom duties on electronic transmission, this work programme did not make much progress; on the contrary, the activity has been discontinued several times.[220]

On 25 January 2019, 75 WTO members, among which are China, the EU and the U.S., adopted a joint statement that re-opened plurilateral negotiations on trade-related aspects of electronic commerce.[221] While not all negotiation positions are public, the EU Proposal for new WTO Disciplines and Commitments Relating to Electronic Commerce is publicly available.[222] According to this proposal, the aim is to negotiate "a comprehensive and ambitious set of WTO disciplines and commitments"[223] for electronic commerce

[215] See for details Irion and Williams (n 18); Susan Ariel Aaronson, 'Data Minefield?: How AI Is Prodding Governments to Rethink Trade in Data' (2018) 11.

[216] Irion and Williams (n 18) 19.

[217] Burri (n 212) 17f.

[218] WTO, Geneva Ministerial Declaration on global electronic commerce, 20 May 1998, WT/MIN(98)/DEC/2.

[219] WTO, Work programme on electronic commerce, 30 September 1998, WT/L/274.

[220] See for background Sacha Wunsch-Vincent, 'Trade Rules for the Digital Age' in Marion Panizzon, Nicole Pohl and Pierre Sauvé (eds), *GATS and the Regulation of International Trade in Services* (Cambridge University Press 2008).

[221] WTO (n 19).

[222] WTO (n 7).

[223] Ibid.

which has come to epitomize digital trade.[224] Meanwhile, it has been reported that Australia, Singapore and Japan have consolidated the proposals into a single document that shows over which issues negotiating parties converge or diverge.[225]

The tabled disciplines and commitments present a mix of eliminating barriers to cross-border digital trade and positive harmonization of domestic rules.[226] The EU proposal backs positive harmonization of electronic contracts, electronic authentication and electronic signatures, among others.[227] Besides, the EU proposal contains new disciplines that relate to software source code, data flows and localization and net neutrality. These:

- Prohibit a member's measures that require the transfer of or access to source code of software, subject to specific derogations;
- Limit members' use of specific data and technology localization measures, subject to a broad exception for members' safeguards to ensure the protection of personal data and privacy; and
- Guarantee open internet access in the sense that members should allow the access, distribution and use of services and applications at the discretion of end-users and their ability to connect devices of their choice to the internet.

3. EU PROPOSAL FOR A SOURCE CODE DISCIPLINE

The ongoing WTO electronic commerce negotiations also take aim at a new discipline on source code of software. The inclusion of source code protection into plurilateral trade rules on electronic commerce is frequently justified as a way to preclude forced technology transfer by parties to such an agreement:

> "Concerns have been raised about the use of registration, certification and approval procedures by government bodies to request, formally or informally, sensitive proprietary information which does not appear to be necessary, or indeed requirements to disclose source code."[228]

Measures that force foreign companies to divulge propriety source code as a condition for market entry or in the context of foreign direct investment can be considered extortionate. Such measure can in particular interfere with business secrets which are often central to business models in high-technology sectors.

a. Proliferation of a source code discipline

Against this background international trade deals are increasingly used to outlaw measures that require access to source code as a condition for market access and/ or foreign direct investment.[229] Table 2 below provides an overview of the source code disciplines in three mega-regional trade agreements without the participation of the EU.

---

[224] Which is arguable a much wider concept Yakovleva and Irion (n 127) 10.

[225] The bracketed draft is not published, see e.g. Iana Dreyer, 'The Basis of an Electronic Commerce WTO Plurilateral Starts Emerging' *Borderlex* (Brussels, 2020) <https://borderlex.eu/2020/08/27/the-basis-of-an-electronic commerce-wto-plurilateral-starts-emerging/>.

[226] See for an overview Gary Clyde Hufbauer and Zhiyao Lucy Lu, 'Global Electronic commerce Talks Stumble on Data Issues , Privacy ', (2019) 19–14 <https://www.piie.com/sites/default/files/documents/pb19-14.pdf>; Rachel F Fefer, 'Internet Regimes and WTO Electronic commerce Negotiations' (2020) Conressional Research Service R46198.

[227] WTO (n 7).

[228] Andrea Andrenelli, Julien Gourdon and Evdokia Moïsé, 'International Technology Transfer Policies' (2019) 222 OECD Trade Policy Papers 4 <https://www.oecd-ilibrary.org/deliver/7103eabf-en.pdf?itemId=%2Fcontent%2Fpaper%2F7103eabf-en&mimeType=pdf>.

[229] See Ines Willemyns, 'Addressing Digital Services in PTAs: Only Convergence in the 11th Hour?' in Rhea Tamara Hoffmann and Markus Krajewski (eds), *Coherence and Divergence in Services Trade Law* (Springer International Publishing 2020) 123 <https://doi.org/10.1007/978-3-030-46955-9_6>.

Table 2: Overview of source code disciplines in mega-regional trade agreements

Trade agreement	Source code discipline
Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP)	**Article 14.17: Source Code** 1. No Party shall require the transfer of, or access to, source code of software owned by a person of another Party, as a condition for the import, distribution, sale or use of such software, or of products containing such software, in its territory. 2. For the purposes of this Article, software subject to paragraph 1 is limited to mass-market software or products containing such software and does not include software used for critical infrastructure. 3. Nothing in this Article shall preclude: (a) the inclusion or implementation of terms and conditions related to the provision of source code in commercially negotiated contracts; or (b) a Party from requiring the modification of source code of software necessary for that software to comply with laws or regulations which are not inconsistent with this Agreement. 4. *[redacted]*
United States-Mexico-Canada Agreement (USMCA)	**Article 19.16: Source Code** 1. No Party shall require the transfer of, or access to, a source code of software owned by a person of another Party, or to an algorithm expressed in that source code, as a condition for the import, distribution, sale or use of that software, or of products containing that software, in its territory. 2. This Article does not preclude a regulatory body or judicial authority of a Party from requiring a person of another Party to preserve and make available the source code of software, or an algorithm expressed in that source code, to the regulatory body for a specific investigation, inspection, examination, enforcement action, or judicial proceeding,[6] subject to safeguards against unauthorized disclosure. [6] This disclosure shall not be construed to negatively affect the software source code's status as a trade secret, if such status is claimed by the trade secret owner.
Regional Comprehensive Economic Partnership (RCEP) Agreement	None

A case in point is the 2018 Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) which is a regional trade agreement between eleven countries.[230] The CPTPP incorporates by reference the original Trans-Pacific Partnership (TPP) signed in 2016 and later abandoned by the incoming US administration.[231] The CPTPP, in its Chapter on Electronic Commerce, prohibits a party to this agreement to require the transfer of, or access to, "source code of mass-market software or products containing such software"[232] as a condition for the import, distribution, sale or use of such software in its territory. The prohibition does not preclude "requiring the modification of source code of software necessary for that software to comply with laws or regulations"[233] which are themselves not inconsistent with this trade agreement.

The chapter on Digital Trade in the 2018 United States-Mexico-Canada Agreement (USMCA)[234] introduces a more ambitious source code discipline. The prohibition to require the transfer of, or access to, source code applies to all software and explicitly covers "an algorithm expressed in that source code".[235] Algorithm is defined as meaning "a defined sequence of steps, taken to solve a problem or obtain a result."[236] The USMCA source code discipline does not preclude "a regulatory body or judicial authority" from requiring access to source code for a specific regulatory investigation or judicial proceeding.[237] Apart from that, a party's violation of the source code discipline can still be justified pursuant to the exceptions foreseen in the USMCA

The Regional Comprehensive Economic Partnership (RCEP) Agreement, signed in November 2020 does not contain a source code discipline in its chapter on Electronic Commerce.[238]

[230] Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) between Australia, Brunei, Canada, Chile, Japan, Malaysia, Mexico, New Zealand, Peru, Singapore, and Vietnam, March 2018.

[231] The text of the TPP is available at <https://www.international.gc.ca/trade-commerce/trade-agreements-accords-commerciaux/agr-acc/tpp-ptp/text-texte/toc-tdm.aspx?lang=eng> (accessed 5 November 2020).

[232] CPTPP Article 14.17.

[233] Ibid.

[234] United States-Mexico-Canada Agreement (USMCA), November 2018. Available at <https://ustr.gov/sites/default/files/files/agreements/FTA/USMCA/Text/19-Digital-Trade.pdf> (accessed 5 November 2020).

[235] USMCA Article 19.6.

[236] USMCA Article 19.1.

[237] Ibid.

[238] The Regional Comprehensive Economic Partnership (RCEP) is a regional free trade agreement between the 10 member states of the Association of Southeast Asian Nations (ASEAN) (i.e. Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, Vietnam) and six partners (i.e. Australia, China, India, Japan, New Zealand and Republic of Korea), November 2020. The text of the RCEP is available at <https://rcepsec.org/legal-text/> (accessed 5 November 2020).

b. The source code discipline in EU trade policy

Though not a party to any of the mega-regional trade agreement mentioned above, also the EU has enlisted a source code discipline in its recent bilateral trade agreements. The EU-Japan Agreement on Economic Partnership, signed in 2018,[239] and the EU-Mexico Agreement in principle, announced in 2018,[240] mark the introduction of a source code discipline in EU external trade policy. The most recent example for a source code discipline can be found in the EU-UK Trade and Cooperation Agreement[241] which aims to cover the relationship after the UK's withdrawal as a member state from the EU. The EU-UK Trade and Cooperation Agreement still has to be approved by the European Parliament.

The text of the source code disciplines in the trade agreement with Japan and in the agreement with the UK is reproduced in Table 3. The language of the source code discipline shows some evolution in the rule-exception-logic that is in the EU-UK Trade and Cooperation Agreement more layered and conditioned. It is clear from the wording that the discipline does not prevent discovery by a court in judicial proceedings or investigations by regulatory bodies or administrative tribunals. Moreover, a party can justify mandating access to software source code in the context of a certification procedure subject to meeting the requirements of the general exceptions and the security exceptions contained in the agreement.

With EU's bilateral trade agreements the intuitive link between the risk of forced technology transfer and the source code discipline is not evident. Neither the EU and its member states nor Japan, Mexico and the UK have so far been implicated with practices that amount to forced technology transfer. Instead, the EU's external trade policy appears to proliferate a template for a chapter on electronic commerce (or digital trade) that is considered best practice (or a gold standard), including, among others, a source code discipline.

[239] See EU-Japan Economic Partnership Agreement (EPA), signed July 2018 and into force since February 2019, Article 8.73. The text of the EU-Japan EPA is available at <http://trade.ec.europa.eu/doclib/press/index.cfm?id=1684> (accessed 5 November 2020).

[240] See EU-Mexico Agreement in principle announced on 21 April 2018, pending ratification <http://trade.ec.europa.eu/doclib/html/156811.htm> (accessed 5 December 2020).

[241] The TCA is already provisionally applied, pending final ratification by the European Parliament and EU member states. European Commission, Draft Trade and Cooperation Agreement Between the European Union and the European Atomic Energy Community, of the One Part, and the United Kingdom of Great Britain and Northern Ireland, of the Other Part, Brussels, 28.12.2020, Article DIGIT.12. <https://ec.europa.eu/info/sites/info/files/brexit_files/info_site/tca-20-12-28.pdf> (accessed 5 December 2020).

Table 3: Overview of source code disciplines in the EU's bilateral trade agreements

Trade agreement	Source code discipline
EU-Japan Economic Partnership Agreement (EPA)	**Article 8.73 Source Code** 1. A Party may not require the transfer of, or access to, source code of software owned by a person of the other Party[1]. Nothing in this paragraph shall prevent the inclusion or implementation of terms and conditions related to the transfer of or granting of access to source code in commercially negotiated contracts, or the voluntary transfer of or granting of access to source code for instance in the context of government procurement. 2. Nothing in this Article shall affect: (a) requirements by a court, administrative tribunal or competition authority to remedy a violation of competition law; (b) requirements by a court, administrative tribunal or administrative authority with respect to the protection and enforcement of intellectual property rights to the extent that source codes are protected by those rights; and (c) [*redacted*]. 3. [*exceptions, redacted*] [1] For greater certainty, "source code of software owned by a person of the other Party" includes source code of software contained in a product.
EU-UK Trade and Cooperation Agreement	**Article DIGIT.12: Transfer of or access to source code** 1. A Party shall not require the transfer of, or access to, the source code of software owned by a natural or legal person of the other Party. 2. For greater certainty: (a) the general exceptions, security exceptions and prudential carve-out referred to in Article DIGIT.4 [Exceptions] apply to measures of a Party adopted or maintained in the context of a certification procedure; and (b) paragraph 1 of this Article does not apply to the voluntary transfer of, or granting of access to, source code on a commercial basis by a natural or legal person of the other Party, such as in the context of a public procurement transaction or a freely negotiated contract. 3. Nothing in this Article shall affect: (a) a requirement by a court or administrative tribunal, or a requirement by a competition authority pursuant to a Party's competition law to prevent or remedy a restriction or a distortion of competition; (b) a requirement by a regulatory body pursuant to a Party's laws or regulations related to the protection of public safety with regard to users online, subject to safeguards against unauthorised disclosure; (c) the protection and enforcement of intellectual property rights; and (d) the right of a Party to take measures in accordance with Article III of the GPA as incorporated by Article PPROC.2 [Incorporation of certain provisions of the GPA and covered procurement] of Title VI [Public procurement] of this Heading.

c. Source code in the WTO electronic commerce negotiations

A discipline on source code of software is part of the ambitious set of new rules on trade-related aspects of electronic commerce currently negotiated by WTO members. There is a history of certain WTO members requesting the protection of software source code in the WTO Work Programme on Electronic Commerce.[242] A number of parties to the negotiations have tabled proposals for a new discipline on source code protection. Canada, for example submitted a proposal that reproduces almost verbatim the USMCA commitment on source code above.[243] The proposals of Japan and the United States are restricted but can be reconstructed based on the CPTPP and the USMCA respectively, to which they are a party, the U.S.-Japan digital trade agreement[244] and other sources.[245] Table 4 below reproduces the texts of proposed commitments on source code of the EU and other key negotiating members insofar as they are public or known otherwise.[246]

Also the EU proposal carries language for a new discipline requiring that members to this agreement "shall not require the transfer of, or access to, the source code of software owned by a natural or juridical person of other members."[247] The proposed discipline on source code protection is subject to certain carve-outs and exceptions. The discipline would be without prejudice to "requirements by a court, administrative tribunal, or by a competition authority to remedy a violation of competition law." The EU proposal moreover carves out commercially negotiated contracts, including public procurement, from the commitment. A source code discipline as is clarified in the EU proposal should be subject to the application of the GATS general exceptions and security exceptions.

[242] Ansgar Koene and others, 'A Governance Framework for Algorithmic Accountability and Transparency' (2019) 67; Willemyns (n 229) 123.

[243] WTO, Joint Statement on Electronic Commerce - Communication from Canada, INF/ECOM/34, 11 June 2019, available at <https://docs.wto.org/dol2fe/Pages/SS/directdoc.aspx?filename=q:/INF/ECOM/34.pdf&Open=True> accessed 5 November 2019.

[244] U.S. Japan digital trade agreement, signed in October 2019, <https://ustr.gov/sites/default/files/files/agreements/japan/Agreement_between_the_United_States_and_Japan_concerning_Digital_Trade.pdf> accessed 11 November 2020.

[245] See e.g. Fefer (n 226); Hufbauer and Lu (n 226).

[246] Restricting access to negotiation positions and documents on the state of play has been criticized for a lack of transparency and deliberative quality. See e.g. Kilic and Avila (n 8).

[247] See Table 2 for the exact wording in the EU proposal, WTO (n 5) para. 2.6.

Table 4: Source code proposals in the WTO electronic commerce negotiations

WTO Member	Textual proposals tabled in the WTO electronic commerce negotiations
European Union[248]	2.6 TRANSFER OR ACCESS TO SOURCE CODE 1. Members shall not require the transfer of, or access to, the source code of software owned by a natural or juridical person of other Members. 2. For greater certainty: (a) the general exception, the security exception […] apply to measures adopted or maintained in the context of a certification procedure; (b) paragraph 1 does not apply to the voluntary transfer of or granting of access to source code on a commercial basis by a natural or juridical person, for instance in the context of a public procurement transaction or a freely negotiated contract. 3. Paragraph 1 is without prejudice to: (a) requirements by a court, administrative tribunal, or by a competition authority to remedy a violation of competition law; (b) the protection and enforcement of intellectual property rights; and (c) [*redacted*].
Canada[249]	ARTICLE 14 Source Code 1. No Party shall require the transfer of, or access to, source code of software owned by a person of another Party, or to an algorithm expressed in that source code, as a condition for the import, distribution, sale or use of that software, or of products containing that software, in its territory. 2. Nothing in this Article precludes a regulatory body or judicial authority of a Party from requiring a person of another Party to preserve and make available the source code of software, or an algorithm expressed in that source code, to the regulatory body for a specific investigation, inspection, examination enforcement action or judicial proceeding,[2] subject to safeguards against unauthorized disclosure. [2] Such disclosure shall not be construed to negatively affect the software source code's status as a trade secret, if such status is claimed by the trade secret owner.
China[250]	Source code not addressed in proposal.
Japan	Confidential proposal.
United States[251]	Confidential proposal, presumably similar to the Canadian proposal.

[248] WTO (n 7).

[249] WTO, Joint Statement on Electronic Commerce - Communication from Canada, INF/ECOM/34, 11 June 2019, available at <https://docs.wto.org/dol2fe/Pages/SS/directdoc.aspx?filename=q:/INF/ECOM/34.pdf&Open=True> accessed 5 November 2019.

[250] WTO, Joint Statement on Electronic Commerce - Communication from China, INF/ECOM/19, 24 April 2019, <https://docs.wto.org/dol2fe/Pages/SS/directdoc.aspx?filename=q:/INF/ECOM/18.pdf&Open=True> accessed 5 November 2020.

[251] WTO, Joint Statement on Electronic Commerce: Communication from the United States, INF/ECOM/23, 26 April 2019 (restricted).

d. What is source code of software?

For determining the scope of a source code discipline the exact meaning of the term source code of software is decisive. The term "source code of software" is neither defined in the EU proposal nor in the other publicly available proposals of other parties to the WTO electronic commerce negotiations. If no definition of source code of software will be provided the rules of treaty interpretation set out in the Vienna Convention would guide the interpretation by the parties to the agreement and ultimately the WTO adjudication bodies. Hereunder a treaty shall be interpreted "in good faith in accordance with the ordinary meaning to be given to the terms of the treaty in their context and in the light of its object and purpose."[252] Following the Vienna Convention "a special meaning shall be given to a term if it is established that the parties so intended".[253]

4.4.1 The ordinary meaning of source code

It follows that in a first step the ordinary meaning of source code needs to be established. The entry in the Oxford English Dictionary provides that inside computing source code means "a program in a source language." Programing languages have been for a long time text-based and human-readable.[254] Computer programs today can involve many thousands of lines of code, collaboratively written by many different programmers. The source code is subsequently converted into machine-readable object code in order to be executed by a computer.[255] Going with the time and technological advancements also programming languages have evolved: older source languages become less relevant and newer source languages are taking over instead.[256] The ordinary meaning of source code is thus not static but dynamically connected to the state-of-the-art of source languages.

Not only business-critical decision-making rules inside the source code of software are protected which is why the source code protection does not align with the protection of business secrets and exceeds the legal protection of trade secrets. All source code is covered even if it fulfils an auxiliary functionality or is incorporated source code that has been written by other organisations or developers. In this context it is worth noting that digitalisation already transforms large realms of public and private spheres into digital artefacts that are consequently represented in source code of software. Any kind of computer programme, software system and even the software architecture of entire online platforms are coded in source language and would does be covered by the ordinary meaning of source code. This means that the scope of a source code discipline is potentially very broad.

4.4.2. Machine learning algorithms are also expressed in source code

In a second step it will be established whether the ordinary meaning of source code also covers computer and/or ML algorithms. 'Hand-coded' computer algorithms are also expressed in source code of software. Human software engineers program the decision-

---

[252] See Article 31 (1) of the Vienna Convention on the law of treaties (with annex), concluded at Vienna on 23 May 1969. Available at <https://treaties.un.org/doc/publication/unts/volume%201155/volume-1155-i-18232-english.pdf> accessed 5 November 2020.

[253] Ibid., Article 31 (4).

[254] See "source code" in Wikipedia. <https://en.wikipedia.org/wiki/Source_code> accessed 9 November 2020, also in Hufbauer and Lu (n 226) 6; Rieke, Bogen and Robinson (n 32) 13.

[255] See "source code" in Wikipedia, ibid., <https://en.wikipedia.org/wiki/Source_code> accessed 9 November 2020, also in Rieke, Bogen and Robinson (n 25) 13.

[256] See "history of source code" in Wikipedia. <https://en.wikipedia.org/wiki/History_of_programming_languages> accessed 9 November 2020.

making rules of computer algorithms 'by hand'.[257] Note in this context that AI research and development is to a fair share based on open source algorithms:

> "Most AI algorithms are shared as open-source code that resides in GitHub, GitLab, or other code repositories. [Deep learning] frameworks, e.g., TensorFlow, PyTorch, Theano, etc. are open source and supported by the largest IT companies such as Google or Facebook."[258]

ML algorithms are increasingly not 'hand-coded' and may no longer rely on text-based programming language that is human-readable. Recent developments see the introduction of machine-generated source code and non-textual code in graphical languages that are used by predictive models.

> "Predictive models tend to be different. They don't take the form of declarative steps, but instead express a statistical relationship between different input and output variables. For example, the "code" for a simple predictive model […] approximates an output variable as a linear function […]."[259]

That does not appear to disqualify visual programming languages and machine-generated code from the ordinary meaning of source code since they are listed as latest developments in programming languages.[260]

Does it matter that the EU proposal for a source code discipline does not include an explicit reference to algorithms? The answer is no because to date an algorithm is commonly expressed in source code using a source language, whether this is hand-coded and text-based or visual and self-learning. Here the Canadian and the U.S. proposals are more straightforward when detailing that also "an algorithm expressed in that source code" would be covered by that discipline.[261] In doing so, the Canadian and the U.S. proposals actually confirm that algorithms are expressed in source code which underscores an interpretation that the source code of software would already cover algorithms.

If at all, the EU proposal for a source code discipline would exclude the more conceptual version of the algorithm before it is translated into source code:

> "Coding thus consists of two key translation challenges centred on producing algorithms. The first is translating a task or problem into a structured formula with an appropriate rule set (pseudo-code). The second is translating this pseudo-code into source code that when compiled will perform the task or solve the problem."[262]

Besides, literature and policy documents in Europe and beyond consistently operate the term source code in connection with AI and algorithmic transparency which would seem odd if algorithms and source code were two different pair of shoes. Below a couple of more references to an algorithm's source code:

---

[257] Mittelstadt and others (n 31) 3.

[258] Blagoj Delipetrev, Chrisa Tsinarakli, and Uroš Kostić. "Historical Evolution of Artificial Intelligence", Publications Office of the European Union, Luxembourg, 2020, doi:10.2760/801580.

[259] Rieke, Bogen and Robinson (n 32) 14.

[260] See "source code" and "history of source code" in Wikipedia.

[261] Even the term algorithm is defined as "a sequence of steps taken to solve a problem or obtain a result." See WTO, Joint Statement on Electronic Commerce - Communication from Canada, INF/ECOM/34, 11 June 2019, Article 1, available at <https://docs.wto.org/dol2fe/Pages/SS/directdoc.aspx?filename=q:/INF/ECOM/34.pdf&Open=True> accessed 5 November 2019.

[262] Kitchin (n 28) 17.

‘‘Algorithmic transparency is not about disclosure of source code as such. It can take different forms, depending on the situation, including meaningful explanation (as required in GDPR), or reporting to the competent authorities … .”[263]

“Legal aspects can also limit certain forms of information disclosure via algorithmic systems. Source codes and hardware designs are often protected as trade secrets.”[264]

“Even if an algorithm’s source code, its full training data set, and its testing data were made transparent, it would still only give a particular snapshot of its functionality.”[265]

Summarizing, based on its ordinary meaning an interpretation of the term “source code of software” includes ML algorithms once they are expressed in a source language.

```
def cnn_model_fn(features, labels, mode):
  """Model function for CNN."""
  # Input Layer
  input_layer = tf.reshape(features, [-1, 28, 28, 1])

  # Convolutional Layer #1
  conv1 = tf.layers.conv2d(
      inputs=input_layer,
      filters=32,
      kernel_size=[5, 5],
      padding="same",
      activation=tf.nn.relu)

  # Pooling Layer #1
  pool1 = tf.layers.max_pooling2d(inputs=conv1, pool_size=[2, 2], strides=2)

  # Convolutional Layer #2 and Pooling Layer #2
  conv2 = tf.layers.conv2d(
      inputs=pool1,
      filters=64,
      kernel_size=[5, 5],
      padding="same",
      activation=tf.nn.relu)
  pool2 = tf.layers.max_pooling2d(inputs=conv2, pool_size=[2, 2], strides=2)

  # Dense Layer
  pool2_flat = tf.reshape(pool2, [-1, 7 * 7 * 64])
  dense = tf.layers.dense(inputs=pool2_flat, units=1024, activation=tf.nn.relu)
  dropout = tf.layers.dropout(
      inputs=dense, rate=0.4, training=mode == learn.ModeKeys.TRAIN)

  # Logits Layer
  logits = tf.layers.dense(inputs=dropout, units=10)

  loss = None
  train_op = None

  # Calculate Loss (for both TRAIN and EVAL modes)
  if mode != learn.ModeKeys.INFER:
    onehot_labels = tf.one_hot(indices=tf.cast(labels, tf.int32), depth=10)
    loss = tf.losses.softmax_cross_entropy(
        onehot_labels=onehot_labels, logits=logits)

  # Configure the Training Op (for TRAIN mode)
  if mode == learn.ModeKeys.TRAIN:
    train_op = tf.contrib.layers.optimize_loss(
        loss=loss,
        global_step=tf.contrib.framework.get_global_step(),
        learning_rate=0.001,
        optimizer="SGD")

  # Generate Predictions
  predictions = {
      "classes": tf.argmax(
          input=logits, axis=1),
      "probabilities": tf.nn.softmax(
          logits, name="softmax_tensor")
  }
```

Figure 1 Example of Tensorflow implementation of a Convolutional neural network in source code[266]

[263] European Commission, A European Approach on Artificial Intelligence, Factsheet, 25 April 2018 <https://ec.europa.eu/commission/presscorner/api/files/document/print/en/memo_18_3363/MEMO_18_3363_EN.pdf>.

[264] Data Ethics Commission (n 3) 170.

[265] Ananny and Crawford (n 28) 982.

[266] Nipun Ramakrishnan (2019), What does AI code look like? <https://www.quora.com/What-does-AI-code-look-like>.

4.4.3 Interfaces to and from an algorithm are expressed in source code

Finally, there is another often overlooked aspect of protecting source code of software as a trade law discipline: the interfaces to and from the algorithm are also expressed in source code. Whenever a software system communicates with users, developers and other third parties this happens via interfaces that are either proprietary or open. APIs are based on industry standards and open protocols that ensure interoperability between different software. Domestic policies that mandate open interface design and interoperability, for example for social networks or electronic commerce websites, would likely violate a future source code discipline and thus must be justifiable inside trade law.

When it comes to ML algorithms also here public-facing APIs and internal APIs are of strategic importance for any meaningful supervision and public scrutiny. A few proposals on algorithmic accountability already highlight the current and future role of interfaces as gateways for auditing algorithms, setting up accountability APIs or experiment with the algorithm in a sandbox setting. Access to these interfaces may turn out crucial to carry out introspection without requiring access to an algorithm's source code using interface audits for example. However, the source code of the interfaces would too be protected as a trade law discipline.

e. What constitutes a violation of the source code discipline?

It is worth noting that the proposed source code discipline differs from the classical trade law disciplines of the GATS. For example the GATS non-discriminatory treatment disciplines, i.e. Most Favored Nation (MFN) and national treatment, take as a point of departure whether a member accords less favourable treatment to a foreign supplier of digital services, either formally or actually, than that afforded to suppliers of any other country or domestic suppliers. It is a comparative discipline that primarily aims to ensure equal treatment between like services and a level playing field for foreign suppliers. The source code discipline by contrast aspires setting a new trade law standard that protects software as such against a party's measure that directly impacts behind-the-border regulations.

The EU proposal for a source code discipline prohibits members to "require the transfer of, or access to, the source code of software owned by a natural or juridical person of other Members." The third paragraph carves out requirements by a court, administrative tribunal, or by a competition authority from the prohibition. The carve-out would typically concern enforcement procedures by domestic regulatory authorities and courts which can request access to source code of software. The second paragraph carves-out situation in which source code of software is voluntarily released on a commercial basis, providing as examples public procurement transaction or a freely negotiated contract.

Since trade law applies to a party's measure, being states and the EU, the main thrust of this source code discipline appears to be general laws and perhaps ad hoc measures. The terms "access" and "transfer" are not defined either, however, inside trade law such terms are commonly ascribed a wide meaning if a measure is found to be trade-restrictive. It follows that a legal measure that effectively divulges software source code would constitute a violation of the proposed discipline. Consequently, prescriptive external auditing of AI systems (i.e. "white box" method) would trigger the scope of the source code discipline. But also legislation that provides for input/ output audits (i.e. "black box" method) would violate the source code discipline if this involves querying an algorithm through software interfaces.

Note that the EU proposal explicitly specifies that the general exceptions and the security exceptions apply to measures in the context of a certification procedure. EU negotiators therefore anticipate that EU certification procedures that require access to source code could be inconsistent with a new source code discipline and therefore in need of a justification.

Finally, a party's measure that mandates the modification of source code is very likely caught by the new discipline judging by the carve-out in the CPTPP source code discipline:

> "Nothing in this Article shall preclude […] a Party from requiring the modification of source code of software necessary for that software to comply with laws or regulations which are not inconsistent with this Agreement."

Following this, other measures that would be inconsistent with the source code discipline are regulations mandating standards on information security, interoperability requirements, or interface design. Should a party's measure be found to be inconsistent with the source code discipline a trade law dispute would move on to the justification stage.

f. Justification of a party's public interest measures

An inconsistent measure can still be justified pursuant to one of the exceptions on grounds of members' security interests (GATS Article XIV bis) or based on the general exceptions for public interest measures (GATS Article XIV). As a rule of thumb, invoking the GATS security exceptions for justifying an inconsistent measure is easier as compared to the more rigorous legal tests required under the GATS general exceptions. It is for the GATS general exceptions to balance trade liberalization objectives with a party's public interest measure and, in doing so, to distinguish legitimate measures from disguised protectionism. Note, however, that reliance on the general exceptions opens a trade law forum according to which it is assessed whether a measure achieves a legitimate public interest objectives in a least trade-restrictive manner. Empirically speaking the intrinsic legal test that is required under the GATS general exceptions, which have been modelled after GATT Article XX, has been very hard to satisfy.[267]

Attempting to justify a measure under GATS Article XIV follows a two-tiered analysis. In a first step, it has to be established that a measure that is found inconsistent with a source code discipline pursues a legitimate general interest objectives that fits the scope of one of the paragraphs of Article XIV. These objectives include measures that are necessary to protect public morals, public order, health and to secure compliance with laws or regulation, including those relating to the prevention of deceptive and fraudulent practices or to deal with the effects of a default on services contracts or the protection of the privacy of individuals in relation to the processing and dissemination of personal data. In a second step, a contested measure needs to satisfy the so-called chapeau of GATS Article XIV requiring that a measure be applied in a manner that does not constitute "arbitrary" or "unjustifiable" discrimination, or a "disguised restriction on trade in services."

The following illustrates how GATS Article XIV(c) would be applied to justify a measure requiring external audits of an algorithmic system, such as would be the case with requiring an independent conformity assessment for AI systems. Meeting the first tier requirement that the measure is designed to secure compliance with laws or regulations which are themselves not inconsistent with the GATS would still be a relatively straightforward exercise. External audits of algorithmic systems certainly contribute to enforcing laws or regulation in the interest of consumer protection, product safety, or anti-discrimination.

The haggling would start over whether external audits are strictly necessary or whether a less trade-restrictive alternative to the measure has been "reasonably available". The less restrictive the measure, and the greater the contribution to the enforcement of public interest, the more likely it is that the measure in question will meet the necessity test. At this stage a measure requiring external audits would be compared against less trade-restrictive measures, such as assessments (i.e. internal audits) and ex post enforcement

[267] See Citizen.org, "Only One of 44 Attempts to Use the GATT Article XX/GATS Article XIV "General Exception" Has Ever Succeeded: Replicating the WTO Exception Construct Will Not Provide for an Effective TPP General Exception", August 2015, available at <www.citizen.org/documents/general-exception.pdf> (accessed 5 November 2020).

by a regulatory authority. This stage gives raise to two issues: first, there is no internationally accepted standard for external audits of AI systems, and, second, transformative AI technology may give rise to new risks for domestic legal systems.

The first issue concerns large variances between countries' approaches to AI, ranging from laissez faire and market-led approaches to fundamental rights' preserving, ethical and trustworthy AI governance. In areas of domestic policy making where there is no internationally accepted standard (yet) defending a high level of protection as compared to less trade-restrictive practices of other countries can be a difficult call. The 2019 OECD Recommendation on Artificial Intelligence, for instance, calls for responsible and trustworthy AI and promotes transparency, explainability and accountability of AI systems.[268] The Recommendation does not make any recourse to the verification and inspection of AI systems, and would thus not lend credibility to a measure requiring external auditing.

As regards the second issue, AI technology is widely perceived as a disruptive technology that can transform all aspects of contemporary life and society, for good or worse. The transnational supply of AI technology affects societies it interacts with and may export risks for fundamental rights, consumer rights, protected values and society to the receiving countries. How possible AI risks for society, democratic institutions or equal treatment of consumers, to name but a few, would be balanced against trade objectives has not yet been tested. Whether, for example, AI's characteristic opacity and scalability, that enables mass-personalization would be recognised as risk-aggravating factors that could swing the necessity test in favour of a justification is not certain.

Should a measure requiring external audits pass the necessity test then the second tier requirements that are in the chapeau of GATS Article XIV must be met. The measure has to be applied consistently without discrimination and discretion has to be exercised reasonably. A measure should not be applied in a discriminatory manner treating any supplier, whether domestic or foreign, the same. The chapeau is an open invitation to seek for inconsistencies in the application of a trade-restrictive measure. A differential application of the measure could for example arise by the requirement to use certified testing centres which are more often located in the EU than abroad. Another reason for finding an inconsistent application can be that a measure requiring external audits affords differential treatment to suppliers from different sectors constituting an "arbitrary" or "unjustifiable" discrimination.

THE SOURCE CODE DISCIPLINE, AS PUT FORWARD BY THE EU, IS ILL-DEFINED AND SIGNIFICANTLY OVERREACHES ITS DECLARED OBJECTIVE TO OUTLAW FORCED TECHNOLOGY TRANSFER.

Following an interpretation of source code also a ML algorithm's source code as well as source code of technical interfaces of an algorithmic system are protected inside the scope of this discipline. A violation of a new discipline that protects the source code of software from a party's measure that request transfer of, and access to, that source code can happen fairly easy with any domestic law that engages with software at a more technical level. Justifying a trade-restrictive measure pursuant to the general exceptions is disproportionately more cumbersome because the legal test and standard of proof can be hard to satisfy, especially in areas of domestic policy where there no international standard and that tackle new digital technologies, such as algorithmic accountability and external audits.

[268] OECD (2019), Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449), adopted by the OECD Council on 22 May 2019 <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

4. RELATIONSHIP WITH COPYRIGHT AND TRADE SECRET PROTECTIONS

In addition, a new source code discipline is a strange guest under the roof of trade-related aspects of electronic commerce. After all, source code of software can be copyright protected and may qualify as a trade secret: both are rights which can be invoked against a country's disclosure requirements.[269] For reasons of consistency source code protection is more appropriately dealt with inside the WTO Intellectual Property (TRIPS) Agreement.[270] Following TRIPS Article 10.1, computer programs, whether in source or object code, qualify for copyright protection. Parties had to align their national laws in order to afford copyrights protection of computer programs in their domestic laws and via its dispute resolution procedures the TRIPS is enforceable.[271]

Next to qualifying for copyright protection, source code can also be protected as a trade secret under the TRIPS agreement. This is confirmed by an explanatory footnote in the Canadian proposal for a WTO source code discipline, which verbally reproduces USMCA Article 19.16, stating "software source code's status as a trade secret" shall not be negatively affected "if such status is claimed by the trade secret owner." The explanatory footnote aims to ensure that source code of software continues to be treated as a trade secret in the case of disclosure in the context of a regulatory investigation, enforcement action or judicial proceeding, requiring safeguards against unauthorized disclosure.

Back to the TRIPS, it has been argued that such a source code discipline would exceed the level of protection afforded to trade secrets under the TRIPS Agreement:

> "This is because Art 39 TRIPS only requires WTO Members to allow the trade secret/confidential information owner to sue someone who obtains/uses it etc. in a dishonest commercial manner." [272]

The language of trade secret protection would actually speak to the concern about forced technology transfer where a government requires transfer of source code for further exploitation in a dishonest commercial manner.

This begs the question what would be the added value of introducing a source code discipline in a prospective agreement on electronic commerce under the roof of the WTO? Often legal certainty is named as an important motivation to inject a source code discipline in order to eliminate, for example, the unwanted practice of forced technology transfer. However, as envisaged a source code discipline creates an additional quasi-proprietary right that, next to copyright protection and trade secret law, shields software source code from being interfered with by domestic measures. However, why source code of software should enjoy triple protection as copyright protected material, trade secret and sui generis software source code is not well founded or explained.

Finally, note the unconditional carve-out from the scope of the proposed source code discipline for the protection and enforcement of intellectual property rights. This serves to exempt legislation requiring disclosure of source code, for example when registering a patent, or situations where there is a conflict over intellectual property rights that can only be resolved by access to the software source code under dispute. The result would be an interlocking system of legal protections at international and national levels that

---

[269] Irion and Williams (n 18); Maggiolino (n 50).

[270] Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), Annex 1C to the Agreement establishing the World Trade Organization, signed in Marrakesh, Morocco on 15 April 1994 <https://www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm>.

[271] Aaron D. Charfoos, How Far Have We Come, and Where Do We Go from Here: The Status of Global Computer Software Protection under the TRIPS Agreement, 22 Nw. J. Int'l L. & Bus. 261 (2001-2002).

[272] Sanya Reid Smith, 'Some Preliminary Implications of WTO Source Code Proposal', vol Third Worl (2017) 3 <https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S001.aspx.>.

shield software source code from interferences, unless it serves the purpose of intellectual property rights.

THE WTO ELECTRONIC COMMERCE NEGOTIATIONS DO NOT MAKE A CONVINCING CASE FOR TOPPING-UP THE PROTECTION OF SOFTWARE SOURCE CODE.

What could not have been achieved using the existing international instruments, e.g. the TRIPS, will not be fixed by an additional layer of protection for software source code under trade law. [273] By contrast, the repercussions for domestic policymaking on digital matters, ranging from accountability, certification, interoperability, portability, to verifiability of digital technologies are potentially vast considering the transformative impact of digitalization and AI technology on all spheres of society and the early stage of this transition we are currently in.

5. HARMONISING CONSUMER PROTECTION WITHIN TRADE LAW

Introducing new disciplines that aim at enhancing consumer trust in electronic commerce transactions cannot offset the negative effects of a new source code discipline on the accountability of transnational digital technologies. Central to the EU proposal are two new disciplines that seek protection of consumers from fraudulent and deceptive commercial practices in electronic commerce as well as from unsolicited commercial communications.[274] The remainder of the EU proposal concerns non-binding standards to require bona fide trading practices, to provide accurate information on the goods or services and the terms of the contract and to grant consumers access to redress. The EU seeks WTO Members' recognition of the importance of cooperation between their consumer protection agencies or other relevant bodies in order to protect consumers.

Better recognition of consumer protection interests in the WTO electronic commerce negotiations has been an important political objective for consumer protection organizations in Europe and beyond.[275] Even if new language on consumer rights will be incorporated into a new WTO agreement on trade-related aspects of electronic commerce, it would not mitigate the multifaceted risks European consumers face from transnational AI technology. One reason is that the proposed consumer protection disciplines would only address a fraction of the body of European consumer rights legislation mapped out in Section IV above. Additionally, as long as parties do not maintain binding rules on AI technology, inside domestic consumer protection law or apart, the harmonising effect for consumer-facing AI technology will be limited even in the space covered by a future WTO electronic commerce agreement.

There are further doubts as to whether the WTO is well positioned to achieve positive harmonisation in the field of cross-border consumer protection. With a few exceptions that are essentially market-making[276]:

---

[273] In 2018 the EU and the US have called the WTO consultation and dispute resolution mechanism to decide on whether China's practices violate the General Agreement on Trade and Tariffs (GATT) or the TRIPS respectively. Even though the issue is not yet resolved, some progress has since been made since China outlawed the compulsory transfer of technology requirements that were targeted by the WTO cases. See e.g. European Commission (2018), "EU steps up WTO action against China's forced technology transfers," Brussels, 20 December 2018 http://trade.ec.europa.eu/doclib/press/index.cfm?id=1963&utm_source=dlvr.it&utm_medium=facebook; Ton Zuijdwijk (2019), "Understanding the Intellectual Property Disputes between China and the United States," <https://www.cigionline.org/articles/understanding-intellectual-property-disputes-between-china-and-united-states>.

[274] Ibid, Articles 2.3 and 2.4.

[275] BEUC, 'WTO E-Commerce Negotiations BEUC Recommendations' (2019) <https://www.beuc.eu/publications/beuc-x-2019-014_wto_e-commerce_negotiations_-_beuc_recommendations.pdf>; Consumers International. 'The Consumer Checklist for an international e-Commerce Deal', 2018 <https://www.consumersinternational.org/media/155222/consumerchecklistforinternationale-commercedeal.pdf>.

[276] E.g. the GATS Annex on Telecommunications or the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS).

> "the WTO presently has limited experience in promoting regulatory convergence on trade-related matters."[277]

The WTO dispute settlement system offers an enforcement mechanism vis-a-vis its members, which are thus states or the EU.

> "If a WTO member violated its regulatory commitment to online consumer protection, affected individuals in another country would need to rely on his or her home state to bring a challenge."[278]

Regardless, the dispute settlement system of the WTO is rarely called to enforce a non-trade related general interest objective, which are classically invoked in an attempt to justify a domestic measure that has been found in violation of WTO law.

A TRADE LAW DISCIPLINE ON SOFTWARE SOURCE CODE PROTECTS ALGORITHMS AND AI SYSTEMS AGAINST MEASURES BY GOVERNMENTS IN THE INTEREST OF ACCOUNTABLE AI

Algorithms expressed in source code would be covered by the scope of the proposed source code discipline which would not only outlaw forced technology transfers but as a bycatch many domestic digital policies that engage with software at a technical level. Laws and regulations that mandate external audits in the interest of algorithmic accountability, even an input/output audit via the interfaces to and from an algorithm, would be a violation of such a new source code discipline. Justifying an inconsistent measure inside trade law can be challenging in the emerging field of AI accountability where countries' approaches vary widely and no international "gold standard" exists. Source code of software moreover qualifies for copyright and trade secret protection; both rights are enforceable under the TRIPS dispute resolution procedures and can be invoked against a country's disclosure requirements.

[277] Ioannis Lianos and others, 'The Global Governance of Online Consumer Protection and E-Commerce Building Trust' (2019) 14. See for a critical account of the welfare and efficiency impacts of regulatory harmonization via trade agreements Dani Rodrik, 'What Do Trade Agreements Really Do?' (2018) 32 Journal of Economic Perspectives 73.

[278] Lianos and others (n 277) 15.

V. SOURCE CODE DISCIPLINE MEETS EU GOVERNANCE OF AI AND CONSUMER RIGHTS

This Section connects the information from the previous Sections on EU policy-formation on AI governance and AI risks anticipated for consumer rights with EU trade law obligations, especially with a view to endorsing a new source code discipline. There is a concern that if not carefully conditioned a new discipline that restricts the transfer of, and access to, source code of software inside trade law could prematurely foreclose policy space for introducing meaningful accountability of AI. What worries pundits and civil society is that a new source code discipline in a WTO agreement could preclude or limit domestic regulations on transparency and external auditing of software systems and computer/ ML algorithms that may actually be harmful to consumer interests in accountable transnational AI technology:

> "… some countries have already linked AI with provisions on cross-border transfer of data and disclosure of source code and algorithms in trade agreements. This would restrict or make difficult the introduction of public supervision of AI and algorithmic decision-making."[279]

> "… there is a risk that under these competitive conditions any regulatory intervention to mandate algorithmic transparency may be interpreted as protectionist interventionism intended to block market access by foreign companies."[280]

Due to its novelty, there is currently no experience with a new trade law discipline protecting software source code and insufficient analysis of its scope, application and effects in practice but also in how far a violation of the discipline can be justified.[281] Understanding how a new source code commitment implicates EU policymaking matters for three reasons:

1. to ensure the internal compatibility of EU policy and its trade law commitments;
2. to initiate a democratic discourse about any trade-offs between source code protection inside trade law and EU governance of AI; and
3. to keep pace with the evolving understanding of risks of AI, including for EU consumer rights, and methods to hold AI systems accountable.

This Section also features several examples which have been modelled after policy documents and research on AI's perceived impact on consumer protection in the Union.[282] In order to activate EU trade law obligations, these cases concern the supply of an AI-powered consumer service by economic operators which operate from outside the EU. Even though these cases are fictitious the featured AI functionalities are not far-fetched and already marketed with consumer products.

---

[279] Trans-Atlantic Consumer Dialogue, 'Resolution on Digital Trade (DIGI 02/189)' (2019) <http://tacd.org/wp-content/uploads/2019/02/TACD-Resolution_digitaltrade_Jan2019_final.pdf>.

[280] Koene and others (n 56) 74.

[281] A handful of policy documents are from academics and civil society actors, see Ansgar Koene, 'Some Implications of WTO Ecommerce Proposals Restricting Access to Algorithms on Algorithmic Transparency', *Paper Presented at the WTO Public Forum, 2-4 October 2018* (2018) <https://ourworldisnotforsale.net/2018/Koene_algorithms.pdf>; Sanya Reid Smith, 'Some Preliminary Implications of WTO Source Code Proposal', Third World Network (2017) <https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S001.aspx.>; Deborah James, 'Digital Trade Rules: A Disastrous New Constitution for the Global Economy, by and for Big Tech' (2020) <https://cepr.net/wp-content/uploads/2020/07/digital-trade-2020-07.pdf>.

[282] See e.g. BEUC, 'Survey: Consumers See Potential of Artificial Intelligence but Raise Serious Concerns' *Press Release* (Brussels, 7 September 2020); Sartor (n 43); Jabłonowska and others (n 129).

1. INTERNAL COMPATIBILITY WITH EU POLICIES

When the EU negotiates an international trade agreement on behalf of its member states, the competent EU institutions have to ensure that an agreement is compatible with internal Union policies and rules.[283] Hence, the EU should only commit inside trade law what is in conformity with EU law and policy in the first place. However, due to the novelty of AI governance there is not yet an internal reference framework against which the compatibility of a source code discipline could be tested. However, even if compatibility with internal Union policies and rules is not yet an issue, there should be awareness about the range of policy options discussed in the Union and foresight about preserving a margin of manoeuvre in the new and dynamic field of AI governance.

Simply put, the rise of AI technology and ADM has spurred calls for the regulation of transparency and accountability for developers and providers of AI technology marketed in the EU. In addition to transparency that offers descriptive information about the functioning of an AI system, the ability to look under the hood of technology is significant too. If the EU commits inside trade law to the protection of software source code, its options to verify or standardize digital technologies are curtailed to exactly what margin of manoeuvre is left under the general exceptions in a given trade agreement. However, policy formulation in the EU and member states could strike a different balance between trade secrets and business interests on the one hand and on the other hand risks for consumers and society.

Table 5 below lists several policy options that are currently discussed in the field of AI governance that would likely be found inconsistent with a source code discipline inside trade law. One example stems from the European Commission's White Paper on AI which proposes a conformity assessment for high risk AI applications before they are marketed inside the EU. Another example comes from the European Commission's legislative proposal for a Digital Services Act which puts forward a new form of qualified data access for "vetted researchers" for the purpose of "conducting research that contributes to the identification and understanding of systemic risks" of very large online platforms.[284] The Digital Services Act proposal foresees that such data access should be realized via dedicated interfaces (APIs) which can be queried by those vetted researchers to conduct public interest research.

The conformity assessment at the second position in Table 5 has been modelled after the European cybersecurity certification listed first. Note that within trade law both measures would take a different path to seek justification: a measure in the interest of cybersecurity could be submitted to the GATS security exceptions whereas justifying AI governance and public interest measures, other than security, would call directly on the GATS general exceptions. As it was explained in the previous Section, a trade law forum, i.e. the WTO dispute resolution mechanisms, would assess whether the trade-conforming disciplines of the GATS general exceptions are satisfied.

[283] Consolidated Version of the Treaty on the Functioning of the European Union, Article 207(3)(2), OJ C 326, 26.10.2012, p. 47 [TFEU] <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A12012E%2FTXT>.

[284] See European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM/2020/825 final) <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608117147218&uri=COM%3A2020%3A825%3AFIN>.

Table 5: Policy options inconsistent with a source code discipline

Measure	Example	Issue
External audit	Regulation (EU) 2019/881, Article 56f.[285]	European cybersecurity certification scheme pertaining to ICT products, services and processes, that can be mandatory under EU or MS law
External audit	European Commission's White Paper on AI[286]	Proposal for requiring conformity assessment of high-risk AI applications in the form of an independent audit by certified testing centres
Data access and scrutiny	European Commission's proposal for a Digital Services Act, Article 31[287]	Very large online platforms would be required to enable access to data for vetted researchers through application programming interfaces in order to study systemic risks
Interface audits	Opinion of the Data Ethics Commission[288]	Proposal to facilitate "always-on" regulatory oversight of algorithmic systems which exhibit a high potential for harm (Level 4) through a live interface with the system
Interface audits	A report by Guillaume Klossa, special adviser to European Commission Vice-President Andrus Ansip[289]	Harness transparency obligations to expose specific APIs in order to create algorithmic sandboxes in relation to digital media platforms
Public record	French Digital Republic Bill, Article 2(I) [290]	Software source code used by the French government is classified as a public record subject to transparency laws

[285] Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), OJ L 151, 7.6.2019, p. 15–69.

[286] European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)' (n 2) 23.

[287] See European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM/2020/825 final) <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608117147218&uri=COM%3A2020%3A825%3AFIN>

[288] Note in this context the recommendation to provide for interface audits by regulatory authorities as of moderate risk categories of the Data Ethics Commission (n 3) 184.

[289] Klossa (n 62) 60.

[290] The Digital Republic Bill (*Loi pour une République Numérique*, n° 2016-1321) of 7 October 2016, Article 2(I), (in French) <https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte>.

Insofar as measures of the EU and member states only target AI technology's highest risk levels they are more likely of being justified under the GATS general exceptions which are geared toward minimizing trade restrictions. Other configurations that require arrangements for interface audits at moderate risk categories of AI applications[291] or prioritize general interest objectives over source code protection could risk falling short of a justification under the GATS general exceptions. Note in this context that the Data Ethics Commission in its Opinion on future AI governance in the EU resolves:

> "… rigid rules of priority, for example a general preference for the protection of business secrets over transparency interests, are not appropriate for the matter concerned."[292]

This is possible since the risk and proportionality assessments performed in EU and member states' legal systems necessary for justifying interventions with software source code and those under trade law do not fully converge.

a. Preserve a crucial margin of manoeuvre

The central findings of this study that the scope of a source code discipline would cover algorithms expressed in source code as well as other software components, which are crucial for the accountability of an AI system, are not trivial for preserving the EU's autonomy to regulate in the context of trade agreements. A future agreement on electronic commerce under the roof of the WTO to which the EU will be a party would implicate how the EU can leverage auditing methods that are important for ensuring accountability and trustworthy AI technology. This concerns for example the policy option to introduce general laws that mandate external audits of the algorithmic source code ("black box" method) and to enable interface audits via dedicated APIs ("white box" method) with the aim to foster accountability, verifiability and trust.

Moreover, owing to the complex and dynamic development of transformative AI technology it is difficult to predict all implications and risks for individual users, democratic institutions and society at large. Several recent developments in the field of algorithmically mediated media and political advertisement on social media platforms, for instance, have dramatically changed the outlook of policymakers at EU and member states levels about the need for regulatory intervention that may lead to instituting more permanent oversight over certain algorithms and online platforms:

> "The recent Facebook/Cambridge Analytica scandal, in which it is alleged that data unlawfully harvested from the Facebook profiles of millions of users were utilized for political micro- targeting in ways that may have perverted the outcome of the US 2016 elections and the Brexit 2016 referendum, reveals not only how readily mass personalization techniques can be exploited and abused but also how serious and damaging their consequences might be for the health and integrity of democratic political orders."[293]

[291] Data Ethics Commission (n 3) 179f.

[292] ibid 188.

[293] Yeung (n 152) 262.

Example from the consultation on the White Paper on AI

The consultation of the White Paper on AI produced one excellent example to illustrate how an EU policy proposal can make a rhetorical turn to a trade-restrictive measure.[294] The European Commission proposes prior conformity assessment for high-risk AI applications in order to verify that a digital technology complies with EU law before it is marketed in the EU. In order to perform an independent conformity assessment certified bodies require access to software source code and audit ML algorithms.

Inside trade law, tying EU market access to prior conformity assessments would certainly affect the cross-border supply of digital services, including AI technology. A look at one submission to the public consultation shows the risk of this measure being framed as a protectionist or trade-restrictive measure:

> "Trade law provisions like the ones the EU and the United States support are important for trade and data-driven innovation as they reduce the risk of parties using concerns over 'cybersecurity' or 'algorithmic transparency' as an excuse to enact requirements that they hand over source code as a condition of market entry market entry, which allows them to pass on this valuable intellectual property to domestic firms."[295]

Note that this submission readily makes the connection between the White Paper on AI and a trade law discipline on software source code in order to disqualify a proposal to introduce independent third-party conformity assessments even for high-risk AI applications. This example serves to underscore how important a good apprehension of the intrinsic relationship between EU policy formation on AI governance and trade law's source code protection is for ensuring EU's autonomy to regulate in international trade deals.

It is crucial to acknowledge that public policy formulation today operates under conditions of uncertainty about possible new and unforeseen risks that may arise from particular and wholesale impacts of AI on protected interests in the EU. Currently, we are witnessing is a regular stream of reports about a potentially faulty, biased or unfair AI system somewhere that keeps us alert about undesirable side-effects of AI technology. Owing to the early stage of exposure to and experience with these technologies our current understanding of adequate safeguards to mitigate the risks and promote trust are necessarily confined to this stage of development.

Likewise, state of the art research into mechanisms that can hold AI and ADM systems accountable is still in its infancy, considering that most literature in this field is from the last decade. Striking are in this context the many contributions of experts and academics from different domains and disciplines compiled in this study who recognize the potential of interfaces (APIs) for accountability and trustworthy AI.[296] In light of this converging statements it seems counterintuitive to commit to a new trade law discipline that makes it harder to engage with AI systems and algorithms via these interfaces. It is very important to provide space for the evolutionary development of transparency and accountability mechanisms that can be expected in the future.

---

[294] Note that such conformity assessment may even be caught as a technical barriers to trade (TBT). See Joshua Paul Meltzer and Cameron F Kerry, 'Cybersecurity and Digital Trade: Getting It Right' (2019) <https://www.brookings.edu/research/cybersecurity-and-digital-trade-getting-it-right/>.

[295] Information Technology & Innovation Foundation (ITIF), Response to the Public Consultation for the European Commission's White Paper on a European Approach to Artificial Intelligence, 12 June 2020, Washington, D.C. <http://www2.datainnovation.org/2020-eu-ai-whitepaper-response.pdf>.

[296] See Section II.2.3 on Interface audit and Section III.1.4 on the Opinion of the Data Ethics Commission.

While the EU can in the future adopt stricter measures to mitigate risks from the cross-border supply of AI technology, once the protection of software source code has entered a prospective plurilateral WTO agreement on trade-related aspects of electronic commerce it will not budge. Foresight and precaution would demand:

> "[t]he adoption of a strong position in trade negotiations to protect regulatory ability to investigate algorithmic systems and hold parties accountable for violations of European laws and human rights."[297]

At a time when transformative AI technology is just starting to take root the EU and other countries should better guard their right to regulate in favour of accountable and trustworthy AI.

b. Public information and democratic debate

There is a stark contrast in the way how the European Commission formulates new policy, for example in the field of AI governance, as compared to its external trade policy. Internal EU policy making passes through several stages involving a policy document, a public consultation, and an impact assessment, before the European Commission submits a legislative proposal to the European Parliament and the Council. Stakeholders can think along during the policy-making process and submit statements to the consultations. Digital policy initiatives, such as the making of the GDPR and currently the Digital Services Act package as well as the AI White Paper, generate a lively public debate on the proposals.[298]

The formulation of EU's external trade policy by contrast is shrouded in mystery because here the European Commission does not volunteer any information about the assessment of the internal compatibility of an agreement with internal Union policies and rules as required under EU primary law. Nowhere does the European Commission, which conducts the WTO electronic commerce negotiations on behalf of the Union and its member states, explain its analysis of the trade law interface with AI governance. There are no studies commissioned by the European Commission that would provide independent advice on the impact of a new source code discipline for ensuring accountable and trustworthy AI technology. Also the Council's negotiation directives do not offer an explicit mandate to negotiate trade in AI and to protect computer and ML algorithms as software source code where this would be detrimental to prospective EU governance of AI.[299]

This study resolves that the source code discipline backed by the EU proposal has tangible repercussions for EU's margin to adopt legislation to hold AI and ADM systems accountable where it formulates requirements that affect source code of software. Its findings are derived from an intrinsic and complex chain of legal interpretations which are more implicit rather than in plain sight. By all means EU decision-makers, member states' governments, political actors, social partners, stakeholders and civil society should have a chance to discuss if they approve of a trade law discipline protecting source code of software in light of its consequences for EU's regulatory autonomy to govern AI. While the details of a source code discipline are still being negotiated,

> "… due care will be required to ensure that such clauses in free trade agreements do not cause problems for accountability and regulatory oversight of algorithmic systems."[300]

---

[297] Koene and others (n 56) 74.

[298] See n 125 above.

[299] Council of European Union, The negotiating directives for the Doha Development Agenda regarding the plurilateral negotiations of rules and commitments on electronic commerce, 8993/19 ADD 1, 20 May 2019 <http://www.consilium.europa.eu/media/39505/st08993-ad01-en19.pdf>.

[300] Koene and others (n 56) 74.

What is more, the making of international trade law does not conform to the basic tenets of democratic representation, open deliberation and inclusiveness that characterize law making and policy formation in the EU and its member states. The negotiations at the WTO are not transparent and negotiations take place behind closed doors. Many parties keep their proposals confidential and the consolidated draft prepared by Australia, Japan and Singapore mid-August is classified:

> "The 91-page e-commerce negotiations stocktake text offers a glimpse of what a potential international plurilateral e-commerce treaty could look like."[301]

Even though the EU proposal underlines that it "supports the open, transparent and inclusive character of these negotiations",[302] this simply is not the ground truth in the ongoing WTO electronic commerce negotiations.

2. ENSURING A HIGH LEVEL OF EU CONSUMER PROTECTION

Section IV on AI risks anticipated for EU consumer rights testifies to the enduring challenges for current regulatory formations to be assertive in digital consumer markets increasingly powered by AI systems and mass-personalisation. EU consumer law's principle of the protection of the weaker party and the precautionary principle would dictate a consumer rights approach to AI governance. The Committee on the Internal Market and Consumer Protection of the European Parliament underscores:

> "the need to look beyond the traditional principles of information and disclosure on which the consumer acquis has been built, as stronger consumer rights and clear limitations regarding the development and use of algorithmic systems will be necessary to ensure technology contributes to improving consumers' lives and evolves in a way that respects fundamental and consumer rights and European values"[303]

If, however, future EU rules on AI governance will not apply to the bulk of consumer-facing AI, AI's characteristic opacity will stand in the way of gaining positive knowledge that an AI system is faulty, biased or unfair. If however EU rules on AI governance that focuses on high-risk AI applications do not include EU consumer rights, even though infringements can affect a large number of EU consumers, enforcement will be difficult. Experts recommend consumer protection authorities to develop synergies with other regulators, such as data protection authorities and competition authorities, and avail themselves of support from technical experts.

a. Harnessing qualified transparency

Consumer markets are among the first to experience mass personalization in advertisement, transactions, and recommendations which call for regulatory supervision as well as monitoring the effects of personalised pricing practices on consumers. Interrogating commercial practices for their compliance with EU consumer rights and anti-discrimination law would require the analysis of real-world data by regulatory authorities and consumer protection organisations.

> "Transparency would likely have to include audits or control of how data-driven and targeting software operates, in order for consumer protection authorities to develop the ability to assess – in-house or perhaps through outsourced expertise

---

[301] Dreyer (n 225).

[302] See WTO (n 5).

[303] European Parliament, 'Opinion of the Committee on the Internal Market and Consumer Protection for the Committee on Legal Affairs with recommendations to the Commission on the framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)', para. 9 <https://www.europarl.europa.eu/doceo/document/IMCO-AD-648496_EN.pdf>.

> – what the combination of algorithms and use of big data sources are leading to, and to discover the use of erroneous data." [304]

Harnessing qualified transparency to its fullest would certainly help to generate the necessary insights while stepping up oversight and enforcement in the highly technical field of AI. Interface audits are a promising instrument in the toolbox of qualified transparency:

> "to carry out what are known as input-output tests, which check, for example, whether an algorithmic system systematically discriminates against groups."[305]

In order to equip consumer protection authorities, among others, with the necessary competences and tools new legislation may be necessary to institute inspections of the inner workings of AI systems and mandate standardized interfaces for carrying out input/output testing, for example to monitor digital services using mass-personalization.

Across regulatory fields, it emerges that our current system of selective enforcement in individual cases after an infringement has occurred lacks teeth in digital consumer markets increasingly powered by AI and ADM systems. Instead there should be more mechanisms that are mandatory upfront (e.g. publishing impact assessments, independent conformity assessments and certification systems) and that are scalable across the EU internal digital market. Ideas include mainstreaming sector-wide enquiries, carrying out empirical research via audit interfaces, ensuring public scrutiny, especially when it comes to population-wide ADM systems, as well as developing empowering consumer technologies.

Enlarging the capacity of civil society players to represent consumer interests is considered an effective way for holding algorithmic systems accountable:

> "This civil law approach has particularly strong market focus and is characterized by swift responses and is therefore, by international standards, very successful. Associations are essentially politically and administratively independent and can therefore advocate, on their own authority and in the common interest of consumers and companies, for competition regulations and consumer rights to be efficiently protected against unfair business practices which are also damaging for consumers."[306]

Private enforcement, however, faces the dilemma that a civil law court can order disclosure, discovery and evidence production only after a complaint has been lodged. Litigants need to substantiate their claim when initiating a legal procedure concerning an infringement of consumer protection law. In the context of AI and ADM systems obtaining the requisite prima facie evidence would require information, which typically only the trader has, for example about the treatment of other individuals which is not easy to obtain from public sources. Counterbalancing the information asymmetry with higher transparency obligations that allow for public scrutiny may be justified in situations of population-wide AI applications, such as in the case of mass-personalized consumer offerings.

---

[304] Koene and others (n 56).

[305] Data Ethics Commission (n 3) 199.

[306] ibid 204.

Case 1: The Virtual Personal Assistant that personalizes prices

The first case study involves AI-powered Virtual Personal Assistants (VPA) used in consumer devices which are diffusing at a spectacular rate throughout Europe. A VPA is a software program that can interact with an end-user in a natural way, answer questions, follow a conversation and accomplish a variety of tasks.[307] Technically speaking, a VPA is an AI-driven conversational agent which can automatically perform a range of private and personal tasks for the end-user.[308]

For the purpose of this case study the VPA is the lead feature of a connected multifunctional household device that is supplied by a leading U.S. online company to consumers in the Union.[309] Alice is an avid German user of her VPA. As a personal and home device it goes without saying that through the VPA much personal and behavioural information about Alice can be obtained.[310]

Through her VPA Alice enquires prices for soda makers and later places an order. When carrying out her request the VPA interacts with a pricing algorithm on an affiliated market place. The VPA brokers Alice's consumer profile that will lead to a personalized price offer for the soda maker. The VPA communicates the price offer to Alice and, when she hesitates with placing an order, the VPA informs Alice about an additional discount.

Alice is not aware that her VPA acts as a broker of her consumer profile, defaults to an affiliated market place and that both the price and the discount she was offered are personalized. She is not aware that her gender, address and socio-economic status as well as her history of returning online purchased products have influenced the personalized price she was offered.[311] Alice is but one household of an estimated 17 million households in the EU using the same VPA.

While price discrimination is permitted under EU law, gender discrimination in consumer contracts is prohibited. Alice does not know that Bob, another user, could order the same soda maker for less because in his case the pricing algorithms calculated a lower willingness to pay. If Alice would have a right to turn off personalization she would be offered a better price for the soda machine.

Monitoring whether price discrimination is based on gender would require data about both Alice and Bob and many more personalized prices and transactions. A sector-wide enquiry into pricing algorithms by consumer protection authorities based on information requests with electronic market places could reveal discriminatory pricing but cannot be sustained throughout. Always-on monitoring of personalized pricing via standardized interfaces would provide a better mechanism for supervision and enforcement.

[307] See European Commission, 'The Rise of Virtual Personal Assistants' (n 186) 2.

[308] See European Commission, 'The Rise of Virtual Personal Assistants' (2018).; Ruhi Sarikaya, 'The Technology Behind Personal Digital Assistants: An overview of the system architecture and key components' (2017) 34(1) IEEE Signal Processing Magazine 67-81.

[309] See for Amzon's Alexa Vladan Joler and Kate Crawford, 'Anatomy of an AI System' <https://anatomyof.ai/img/ai-anatomy-publication.pdf>.

[310] European Commission, 'Antitrust: Commission Launches Sector Inquiry into the Consumer Internet of Things (IoT) (IP/20/1326)' *Press Release* (Brussels, 16 July 2020) <https://ec.europa.eu/commission/presscorner/detail/en/IP_20_1326>.

[311] Rebecca Smithers, "Boots revises cost of two products over accusations of sexist pricing," The Guardian, 2 February 2016 <https://www.theguardian.com/business/2016/feb/02/boots-alters-prices-accusations-of-sexist-pricing>.

Through the lens of trade law, the envisaged source code discipline recognises qualified transparency that tends to be applied on a case-by-case basis. As has been explained in Section IV, a source code discipline is without prejudice to requirements by a court, regulatory body or competition authority which use their powers in individual cases in order to investigate whether an infringement has occurred. Legislation that mandates independent conformity assessments, certification schemes or standardised interfaces for the purpose of carrying out regulatory supervision or public interest research would be inconsistent with the protection of software source code inside trade law unless it can be justified under the GATS general exceptions.

b. Public scrutiny of AI systems

The scalability and mass personalization techniques that AI technology facilitates may require different regulatory formations than the ones adopted in the industrial age. AI governance that primarily rests on enforcement in individual cases after an infringement has occurred is probably not agile enough to hold fast moving and scalable AI applications accountable.

> "Relying on regulators to perform all of this research is not advisable, since regulators are capacity-constrained and often lack much of the essential expertise needed to oversee this vast and highly technical field." [312]

Social scientists argue that algorithmic governance would benefit from "regulation towards auditability"[313] that privileges public scrutiny over internal audits. In fact, many noteworthy reports and news about faulty, biased or unfair outcomes of AI technology stem from consumer rights organisations, investigative journalists, digital advocacy groups, and researchers.[314]

> "By mobilizing academics, media, civil society or other independent researchers, policymakers can bring a wealth of expertise and research capacity to bear on urgent regulatory issues – a wealth that no reasonable amount of regulatory funding can match."[315]

Compared to conventional enforcement mechanisms public scrutiny is more agile and ranges from basic observation of an AI system to more sophisticated types of scrutiny, such as public interface audits ("black box" method).[316]

[312] Ausloos, Leerssen and ten Thije (n 80) 15.

[313] Sandvig and others (n 54); Rieder and Hofmann (n 62) 22.

[314] See e.g. Rieke, Bogen and Robinson (n 32); Chiusi and others (n 9); AlgorithmWatch and Bertelsmann Stiftung (n 9); Sandvig and others (n 54); Bodo and others (n 54).

[315] Ausloos, Leerssen and ten Thije (n 80) 15.

[316] Rieke, Bogen and Robinson (n 29).

Case 2: **A booking platform that analyses personality traits**

The second example considers an online booking provider based in the U.S. which connects via its online platform supply and demand for short stay rentals in the Union. EU consumers can create a profile on the online platforms, browse the accommodation options per destination and make reservations. After a promising trial period in the U.S. the booking platform rolls out a new AI-driven software on its platform to profile users and reduce negative business impact of “undesirable” users.

The new software automatically searches social media and news to obtain intelligence about an individual user. The available information, consisting of personal data, photos, reactions, friends and groups etc., are analysed for behavioural and personality traits.[317] The software automatically classifies user for traits of neuroticism, involvement in crimes, narcissism, Machiavellianism, psychopathy, negative language, involvement in pornography, negative news stories, drugs and alcohol abuse.

In our fictitious scenario Bob is a frequent user of the booking platform. He works for a bespoke Belgian wine retailer for which he regularly travels the European wine regions. Bob is a member of various associations of sommeliers, a fervent fan of Greek mythology and frequently he posts on social media about his passion. Lately, Bob finds it somewhat more cumbersome to be offered attractive rentals via the accommodation booking platform.

What Bob does not know is that the aforementioned algorithm has assigned a higher risk score to him based on traits in three categories: psychopathy, negative language in social media, and substance abuse. What happened is that Bob’s passion for Greek mythology misaligned with the algorithm which could not place all the bizarre and disturbing postings on social media. Also the prominence of alcoholic beverages on the images that Bob posts influenced the risks score for substance abuse.

If suspecting a mistake and being informed about his GDPR data subject rights, Bob can request access to his personal data and object against the use of the automated profiling with the booking provider. Bob manages to obtain human intervention and the customer service agent eliminates the risks score for substance abuse but not his risk score for psychopathy and negative language. That the algorithm assigned a negative risk score on more grounds does not surface during the customer service call.

Input/ output audits can help detect how the AI-driven software responds to different customer profiles. A Dutch digital rights group studies the online booking platform with the support of 70 volunteers and compares the results for identical requests. The group discovers that the algorithm is biased and unfair. An auditing API would make empirical research on this algorithm much more effective and help detect faulty, biased, or unfair AI.

[317] Mark Blunden, “Booker beware: Airbnb can scan your online life to see if you’re a suitable guest”, Evening Standard, 3 January 2020 <https://www.standard.co.uk/tech/airbnb-software-scan-online-life-suitable-guest-a4325551.html>.

Unlocking public scrutiny could be an important means to counterbalance the grassing information asymmetry in relation to those who control AI technology, especially in the context of population-wide AI systems linked with systemic risks. Current proposals are primarily linked to specific sectors where AI systems significantly impact on societal interests, such as algorithms that can influence public opinion or those which can cause major welfare effects for the population.[318] In the future new situations may arise where additional society interests or public welfare issues are at stake which call for enhanced public scrutiny of the responsible AI systems.

Consumers are bound to experience the many benefits and potential risks of AI technology first-hand. Detecting faulty, biased or unfair outcomes of AI technology in consumer markets would benefit from "regulation towards auditability" that opens pathways for vetted consumer protection organisations and other public interest groups to perform input/output testing and interface audits in justified cases of public interest research subject to safeguards for business and trade secret protection.

From the perspective of a new trade law discipline that protects against measures that require access to software source code introducing domestic regulation that would selectively enable new forms of public scrutiny will hardly be possible. While designed for recognising enforcement by regulatory authorities and the courts the discipline would shield the interfaces (or APIs) of an AI system from public scrutiny. In its current form a source code discipline and the attendant exceptions would tolerate strictly risk-based measures that are enforced in individual investigations and procedures to govern scalable AI technologies.

c. AI for consumer empowerment

Consumers will need assistance and support to navigate digital consumer markets and assert their individual rights under EU and member states' laws. AI technologies holds great potential for empowering consumers which may help alleviate information and power imbalances and assist consumers in their daily interactions with a plethora of AI applications of economic operators.

> "Here is where AI could play a crucial role: that of driving technologies able to empower consumers and their organizations, by supporting consumers in safeguarding their privacy, defending their rights, protecting them from unfair practices. A real and effective counter-power of consumers against producers and intermediaries needs to be brought about, not to build instruments that represent alternatives to the law, but to overcome the difficulties for consumers and regulatory agencies in enforcing the law."[319]

Consider the innovative idea of "AI Guardians"[320] which connote a digital service that operates strictly in the interest of individual consumers. The idea of AI Guardians takes inspiration from existing privacy management systems and Virtual Personal Assistants (VPA). Contrary to a VPA that is provided as part of a digital platform operator, AI Guardians should not be affiliated to an economic operator but be an independent, non-for-profit endeavour that serves the best interest of individual consumers.

---

[318] Data Ethics Commission (n 3) 187, 210. See for current proposals in relation to social media's grassing problems with mis- and disinformation, news recommender systems and online political advertisements: European Commission, 'EU Code of Practice on Disinformation' <https://ec.europa.eu/digital-single-market/en/news/code-practicedisinformation> accessed 5 November 2020; Council of Europe Committee of Ministers to Member States, Recommendation CM/Rec (2018)1 on Media Pluralism and Transparency of Media Ownership (Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies), para 2.5; Klossa (n 62).

[319] Lippi and others (n 198).

[320] Lippi and others (n 195); Franco Zambonelli and others, 'Algorithmic Governance in Smart Cities: The Conundrum and the Potential of Pervasive Computing Solutions' (2018) 37 IEEE Technology and Society Magazine 80 <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8371566>.

Case 3: **Smart washing machines that optimise warranty liability**

The third case focuses on smart household appliances. Consider a South Korean manufacturer who is using AI to manage warranty liability of its washing machines which are marketed in the EU.[321] The ML algorithm analyses individual devices' feeds while remotely controlling washing machines' functionalities. During the warranty period suspected risks of water leakages are proactively dealt with by issuing a service alert and contacting the customer offering gratuitous maintenance.

With regards to other, less obvious defects, the algorithm optimizes warranty liability by selectively delaying fault messages for those washing machines of which the legal warranty period is close to expiring.[322] While doing so the algorithm factors in which customers are more likely to apply for warranty and who are not. The algorithm thus differentiates the level of customer support offered based on its predictions, thereby treating customers differently.

From a large customer base in Europe, Alice and Bob both own the same high-end model of the smart washing machine. When Alice's washing machine signals malfunctioning she receives full customer support as part of her warranty claim. Bob is less fortunate because his washing machine only signalled the malfunctioning shortly after the two year warranty period had expired.

Alice and Bob have consumer rights under the EU Consumer Rights Directive (Directive 2011/83/EU) but Bob will not be able to prove that the smart washing machine outsmarted the warranty liability period of two years. The practice would be hard to track or detect because it differentiates from customer to customer. Only long-term monitoring or an audit would reveal that the manufacturer' algorithm discriminates between customers in order to optimize its warranty liability.

AI Guardians can perform functionalities, such as:

- managing privacy and data protections preferences and settings;
- analysing terms of service and use policies, protecting from unfair practices;
- managing digital transactions, digital records and electronic signatures;
- logging of interactions with AI systems, descriptive information of ADM and monitoring legal requirements;
- registering individuals requests for data subjects rights, consumer rights and fundamental rights, e.g. the request for access to personal data, human review, explanations;
- assisting with filing complaints with regulatory authorities; and
- supporting sector-wide inquiries and public interest campaigns with anonymized data to hold AI systems accountable.

The idea for developing AI Guardians matters because in order for consumer empowerment technologies several requirements need to be put in place that support interopera-

[321] See e.g. Prasat Shyam, „New technologies and data can lead to more proactive warranty management", Capgemini Blog, 15 Mai 2020 <https://www.capgemini.com/us-en/2020/05/reduce-warranty-liabilities-with-artificial-intelligence/>

[322] See e.g. Consumers International, „Build to Fail: Is Planned Obsolescence Really Happening?", Blog [no date] <https://www.consumersinternational.org/news-resources/blog/posts/built-to-fail-is-planned-obsolescence-really-happening/>; Samuel Gibbs, "Apple and Samsung fined for deliberately slowing down phones", The Guardian, 24 October 2018 <https://www.theguardian.com/technology/2018/oct/24/apple-samsung-fined-for-slowing-down-phones>.

bility, standardized interfaces, open protocols, portability and machine-readable information processing. Others too see the potential of accountability interfaces or auditing APIs:

> "One possible approach is to provide individuals or third party auditors with access to "auditing APIs," which allow users to request counterfactual explanations from the service provider, and perhaps compute them directly via the API."[323]

Trade law would be agnostic to the development of AI Guardians until the point that a measure requires that suppliers of digital services that operate in the Union ensure that their services meet the requirements of the software architecture of consumer empowerment technologies. Because this would mandate modifications of proprietary software source code in violation of the envisioned source code discipline it would spark resistance by suppliers of digital services and governments of third countries which could challenge the measure as a trade restriction in a dispute settlement procedure. It is impossible to predict whether it would be possible to defend a domestic measure that puts into place the conditions for such an innovative AI Guardian to discharge its functionalities in the interest of consumers.

EU MEASURES TO ENSURE ACCOUNTABILITY OF AI SYSTEMS AND A NEW TRADE LAW DISCIPLINE ON SOURCE CODE OF SOFTWARE DO NOT ALIGN

The need to ensure the internal compatibility of EU policy and rules with its trade law commitments should apply foresight and precaution to guard a sufficient margin of manoeuvre that will be necessary to respond to the evolving risks of AI technology and to ensure a high level of consumer protection in the Union. Promising accountability mechanisms that require interventions with or modifications of software source code would be inconsistent with a source code discipline and in need of a justification under trade law. In light of the implications and the early stage of AI development, it would be imperative to initiate an inclusive and democratic discourse about any trade-offs between EU governance of AI and source code protection inside trade law.

Consumers are bound to experience the many benefits and potential risks of AI technology first-hand. Defending consumer rights in digital consumer markets increasingly powered by AI and ADM systems may require more agile and scalable regulatory formations in addition to our current system of enforcement in individual cases after an infringement has occurred. Where appropriate consumer protection should be able to harness collective redress, public scrutiny of population-wide AI systems, and innovative consumer empowerment technologies without being constraint by a trade law discipline on software source code.

[323] Wachter and others (n 74) 882.

CONCLUSIONS AND RECOMMENDATIONS

European consumers already experience first-hand the many benefits of AI applications but also the potential risks of encountering faulty, biased or unfair AI. Digital consumer markets rapidly adopt AI technology which also enables mass-personalisation of online advertisements, content, recommendations, transactions and also prices. The digital global ecosystem allows for cross-border trade in AI and the rapid spread of AI services to EU consumers. This should however not affect the high level of consumer protection guaranteed in the EU and the principle of the protection of the weaker party in consumer law

This study analyses possible trade-offs between the ambition to ensure accountability of AI and a high level of consumer protection and a new trade law clause on software source code. In the ongoing WTO electronic commerce negotiations, the EU backs the introduction of a trade law clause which prohibits a party's measure requiring transfer of, or access to, the source code of software subject to certain exceptions. There is currently no experience with a trade law clause on source code and insufficient analysis of its scope, application and effects on a party's autonomy to regulate. What worries experts and rights advocates is that – if not carefully conditioned – the source code clause could prevent future EU regulation of AI in order to hold transnational AI technology accountable that may be harmful to consumer interests.

> The central finding of this study is that such a source code clause being currently negotiated in plurilateral trade talks for a WTO agreement on electronic commerce, would restrict the EU's right to regulate in the field of AI governance in several important ways. This may be surprising given that EU trade policy documents make no reference to AI technology, only to electronic commerce, and that no direct link has been made between the protection of software source code and computer or machine learning (ML) algorithms. Nonetheless, these findings, which are rather inferred than in plain sight, are derived from a careful interpretation of the source code clause inside trade law.

In order to form a comprehensive understanding and draw these conclusions the analysis was carried out in several stages:

Section I reviews our current state of knowledge about transparency and accountability of AI. It shows that a modular approach to algorithmic transparency is needed which combines different requirements, ranging from information duties, qualified transparency for public authorities and domestic courts to facilitating external audits and public scrutiny in justified cases. Central to accountability, verifiability, and trust in AI are methods to audit algorithms and AI systems. Currently, input/output audits ("black box" method) are used more frequently than auditing an ML algorithm's source code ("white box" method). It emerges that technical interfaces (public-facing or internal APIs) of AI systems are important gateways for ensuring accountable AI.

In Section II the current landscape of policy options for EU legislation on AI governance is surveyed. The European Commission's White Paper on AI foresees new regulation for high-risk AI system that would also apply to economic operators in third countries providing AI-enabled products or services in the EU. Following the Opinion of Germany's Data Ethics Commission more should be achieved, such as regulating AI systems as of moderate risk levels, the publication of ex ante impact assessments, enabling qualified transparency with the help of standardized interfaces to carry out input/output audits and harnessing public interest research in justified cases.

Section III turns to EU consumer protection law and the anticipated risks of AI for consumer rights. It finds that enforcement faces an uphill battle to assert EU consumer rights, due to AI's characteristic opacity and the limited capacities of regulatory authorities to carry out investigations into these technologies. Proposals to overcome these challenges

include alleviating the burden of proof in litigation, stepping up regulatory enforcement capacity and technical expertise, as well as leveraging collective redress and public scrutiny of AI systems.

Turning next to EU's trade law commitments, Section IV resolves that the GATS already applies to cross-border trade in AI-powered digital services. What is more, following a careful interpretation, computer and ML algorithms are expressed in source code and would thus be covered by the scope of a trade law clause on software source code. The broad scope of the source code clause would not only outlaw forced technology transfers but a variety of measures that can hold an AI system accountable would be inconsistent with this discipline.

The source code clause is without prejudice to requirements by a domestic court, administrative tribunal, or by a competition authority, all of which typically take place in individual proceedings after an infringement occurred. General law and regulations, however, requiring access to software source code in the interest of accountability of AI would be inconsistent with the trade law clause. The trade law clause would condition auditing at the level of source code ("white box" method) but also auditing of inputs and outputs of an AI system via its interfaces ("black box" methods) to the margin of regulatory autonomy left under the GATS general exceptions. The GATS general exceptions are geared to minimize the trade-restrictive effect of a measure and have been – empirically speaking – not a robust defence.

In Section V the different strands of the argument are brought together. According to Article 207(3) TFEU, the Council and the European Commission are responsible for ensuring that the negotiated trade agreements are compatible with internal Union policies and rules. The plurilateral negotiations for a WTO agreement on electronic commerce and EU policy formation on AI regulation are proceeding in parallel so that there is currently no EU reference framework with which to ensure compatibility.

Several policy options that are currently discussed in the field of AI governance risk being inconsistent with a clause on source code, unless they can be justified under the general exceptions inside trade law. For example:

- The White Paper on AI proposes the introduction of prior conformity assessment of high-risk AI applications by certified testing centres;[324]
- Germany's Data Ethics Commission recommends "always-on" regulatory oversight of algorithmic systems with a high potential for harm through a live interface;[325] or
- The Digital Services Act proposal requires very large online platforms to enable vetted researchers to study systemic risks by accessing data via interfaces (APIs).[326]

What is already critical now is the strategic importance of interfaces (public-facing APIs and internal APIs) for ensuring accountable and trustworthy AI.[327] Experts and academics from different domains and disciplines highlight the crucial role of interfaces as gateways for auditing algorithms (without requiring access to an algorithm's source code), setting up accountability APIs or experiment with a ML algorithm in a sandbox setting. In light of this, committing to a trade law clause that would make it harder to engage with AI systems via these interfaces or mandate standardized interfaces in the interest of auditability and accountability is counterintuitive.

---

[324] See European Commission, 'White Paper on Artificial Intelligence - A European Approach to Excellence and Trust (COM(2020) 65 Final)' (n 2) 23.

[325] See Data Ethics Commission (n 3) 184.

[326] See European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM/2020/825 final) <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608117147218&uri=COM%3A2020%3A825%3AFIN>

[327] See Section II.2.3 on Interface audit and Section III.1.4 on the Opinion of the Data Ethics Commission.

Another area for future conflict between EU policy and trade law arises where a high level of consumer protection calls for safeguards against anticipated risks of AI technology in digital consumer markets. Adequate safeguards are especially important to alleviate the risks from the cross-border supply of AI-powered services to consumers in the Union by operators from outside the EU. Affirming a high level of consumer protection in the presence of cross-border trade in consumer-facing AI is fairly dependent on ensuring solid accountability of AI systems.

A recurring theme that negatively affects consumer rights across the board is the information asymmetry given the AI systems' characteristic opacity and the associated difficulty of satisfying the burden of proof that an AI system is faulty, biased or unfair in the first place. Defending consumer rights in digital consumer markets increasingly powered by AI requires more agile and scalable regulatory provisions in addition to the current system of ex post enforcement. Consumer protection should harness collective redress, public scrutiny of population-wide AI systems where appropriate, and innovative consumer empowerment technologies without being constrained from the outset by a trade law clause on software source code.

From the perspective of EU consumer protection law and the principle of protecting the weaker party in consumer law, adding an additional layer of protection for source code of software, as the trade law clause on source code does, removes AI systems further from instituting effective accountability and enforcement. Monitoring the effects of AI systems in digital consumer markets would instead benefit from regulation towards auditability, including the ability to mandate external audits and to require standardised interfaces through which input/output audits can be carried out. The scope of a source code clause by contrast would not only cover computer and ML algorithm but also protect the source code of interfaces of an AI system that are indispensable for audits using "white box" methods.

The process of digitalization leads to more and more digital artefacts and transformative AI technology may give rise to new risks for individuals and society. The source code clause appears too broad for domestic digital policies that need to build on interoperability, accountability, and verifiability of digital technologies. If there is a trade-off between EU governance of AI and source code protection inside trade law this should be resolved in a way that respects fundamental and consumer rights and European values. The trade-off should be put to democratic scrutiny and discussion that characterises EU rule-making.

IN TERMS OF POLICY RECOMMENDATIONS THE EU HAS TWO OPTIONS:

1. The European Commission should clarify the impact of the source code clause on EU digital policies, in particular consumer rights, and meanwhile give up on this trade law clause since software source code still enjoys copyright and trade secret protection; or

2. The European Commission should limit the trade law clause on source code of software to:
   a. the situation of forced technology transfers for dishonest commercial practices, or
   b. carve out measures on algorithmic accountability from the scope.

This would be prudent and provide time to develop robust domestic policy as well as international standards on accountable AI. See for proposed language below:

Option 2.a.:

TRANSFER OR ACCESS TO SOURCE CODE

1. Members shall not require the transfer of, or access to, the source code of software owned by a natural or juridical person of other Members with the purpose of re/using it in a dishonest commercial manner.

Option 2.b.:

TRANSFER OR ACCESS TO SOURCE CODE

1. [*redacted*].

2. [*redacted*].

3. Nothing in this Article shall preclude:

(a) a Party from requiring auditing or verification of an algorithm expressed in source code of software that contributes to secure compliance with laws or regulations [which are not inconsistent with this Agreement] and subject to safeguards against unauthorized disclosure of algorithm expressed in source code of software.

(b) a Party from requiring the modification of source code of software necessary for that software to comply with laws or regulations [which are not inconsistent with this Agreement].

3. [*redacted*].