BLAVATNIK SCHOOL OF GOVERNMENT

POLICY BRIEF

PROVIDING ACCESS TO THE LATEST POLICY-RELEVANT RESEARCH

AI GOVERNANCE AND THE FUTURE OF DIGITAL TRADE POLICY

OPTIONS FOR THE UK

Emily Jones (University of Oxford)
Philippa Collins (University of Bristol),
Henning Grosse Ruse-Khan (University of Cambridge)
Albert Sanchez-Graells (University of Bristol)
Kristina Irion (University of Amsterdam)
Cosmina Dorobantu (Alan Turing Institute)
Burcu Kilic (Centre for International Governance Innovation)
Daria Onitiu (University of Oxford)

9 October 2024

EXECUTIVE SUMMARY

Data and digital technologies play an increasingly important role in the global economy and UK firms are increasingly exporting and importing AI products and services. The UK and other governments have been updating trade agreements to capitalise on the opportunities but also address the challenges that arise from the rapid growth in cross-border movements of data and digital technologies. This Policy Brief examines one controversial element of recent trade agreements: provisions on source code non-disclosure. The Brief explains why these provisions on source code pose challenges for the effective regulation of algorithmic and AI systems, where access to source code is needed for essential conformity, assurance and audit checks. We propose ways to revise trade agreements to ensure that the legitimate concerns of businesses seeking confidentiality of their source code are taken into account, whilst also ensuring that the UK is able to introduce effective regulation of AI and algorithmic systems, creating effective accountability mechanisms now and in the future.

The Brief investigates the source code provisions and the ambiguities and restrictions they create for regulating algorithmic and AI systems. We recommend a careful review of the UK's approach to negotiating source code provisions should be undertaken to assess whether there is sufficient evidence to support their continued inclusion in trade agreements and to explore alternative mechanisms for meeting policy objectives. This review must encourage the effective participation of a range of government departments, stakeholders and expertise, including from academia and research institutions, NGOs and citizen's rights organisations, consumer groups, trade unions as well as representatives of the business community. The review needs to also cover parallel concerns regarding emerging cryptography provisions.

Although AI and algorithmic systems bring opportunities, we highlight the need for effective government regulation and oversight of these systems to guard against the individual and systemic risks that they pose. Whilst there are an increasing number of regulatory measures being introduced nationally and internationally, it is not yet possible to determine what effective auditing and evaluation models for complex AI systems may look like *or* the level of disclosure that will be required from AI developers. The interaction between national regulatory measures in this area and the effect of trade rules has been under-investigated. This Brief aims to provide an analysis regarding precisely this question and, crucially, flags the need to avoid consolidation of approaches that could restrict effective regulation.

Source code provisions are included in numerous digital trade chapters and digital economy agreements. They provide that governments signing the agreement will not require 'the transfer of, or access to, source code of software' that is owned by a business originating in their trading partner(s) as a condition for the import, distribution, sale or use of that software. This rule is subject to a range of exceptions across the text of the agreements. We observe that there is growing tension between the core commitment not to mandate disclosure, paired with the narrow scope of the exceptions, and the need for governments to mandate such disclosure as an integral

part of an effective regime for governing algorithmic and AI systems and other emerging technologies.

Difficulties exist at both the "rule" and "exception" level. There are ambiguities regarding the scope of the prohibition on disclosure requirements, and how far it extends to various components of an AI system (for example, training data, data sets, algorithms and Application Programming Interfaces). This creates uncertainty for all parties and may limit the scope of auditing exercises. With regard to the exceptions, we argue that they create a number of difficulties for the UK as it charts a course to effective AI regulation, auditing and accountability. We discuss, amongst other issues, the definition of 'conformity assessment bodies', the need for pre-market auditing of AI systems involving a range of non-state actors, challenges in managing access to source code in the context of public procurement transactions, and potential inconsistencies between the source code provisions and intellectual property law.

In response to these concerns, we propose two alternative ways forward: either the removal of the source code provisions, combined with a modification of existing Free Trade Agreement rules on 'trade secrets' to protect the undisclosed information of AI and other technology developers, or alternatively, a substantial modification of the drafting of source code provisions. Whilst the authors would opt for the former route, in the event the UK government opts for the latter course of action, we make several recommendations:

- expressly and narrowly define 'source code of software' and, in so doing, clarify that Application Programming Interfaces are out of scope;
- reaffirm that existing fair dealing and public interest exceptions in domestic IP law and trade secrets law continue to apply;
- adopt a functional approach to the exemption for conformity assessments (and other audit or evaluation measures as needed);
- allow for a wider range of methods for the audit and evaluation of AI and algorithmic systems, particularly those involving non-state actors (including firms, consumers, workers and researchers);
- ensure that the source code provisions allow for the introduction of government measures requiring the disclosure of information about algorithmic and AI systems to protect the rights and interests of private individuals (firms, workers, and consumers), or their representatives on their behalf, in accountability settings beyond the regulatory and judicial context that is currently mentioned.
- to avert the need to continually update the exceptions to the source code provision to keep pace with government regulation of fast-moving technologies, incorporate a broader regulatory carve out, taking inspiration from the precedent the EU has set in carving out regulatory measures for personal data protection in its digital trade agreements.

CONTENTS

Executive Summary **2**

Contents **4**

Policy Brief **5**

Context 5

Summary 6

The importance of opening the 'black box' of AI systems 7

The ramifications of source code provisions in trade agreements for AI regulation and auditing 10

Ambiguities regarding the scope of source code provisions 13

Narrow scope of exceptions 14

Source code provisions: two options for the future 18

Conclusion 20

Authors 20

POLICY BRIEF

Context

1. This Policy Brief draws on a written response submitted to the House of Lords' International Agreements Committee Inquiry on Data and Digital Trade on 1st October 2024.[1] The growth of the digital economy fundamentally challenges our understandings of how international trade operates, and raises questions about how to adapt existing trade rules and market access commitments to a digital era. The inquiry seeks to understand how developments in digital trade and the digitisation of trade should be reflected in agreements the UK Government negotiates and signs.

2. This Policy Brief is informed by discussions held during a workshop hosted at the Blavatnik School of Government, University of Oxford, by Emily Jones and Philippa Collins on 9-10 September 2024. The participants reflected a range of expertise, including several legal fields (international intellectual property law, AI regulation, workplace and consumer rights, data protection, public procurement), trade policy experts from within academia and think tanks, experts in the field of AI accountability, and trade practitioners.

3. This Brief focuses on the increasing need for better regulation of algorithmic and AI systems, and the tensions created by existing conditions to protect source code within trade agreements. It focuses on the following questions posed by the Inquiry:

    - *How do you think the government should balance issues such as the right to regulate to protect data privacy or to access source code, with commitments in treaties protecting free flows of data or intellectual property of software developers? What has its approach been to date and how do you think it should approach these issues in future?*

    - *How effective would you say stakeholder engagement has been in the development and implementation of digital trade agreements, or in the digital provisions of international agreements?*

[1] https://committees.parliament.uk/work/8432/data-and-digital-trade/publications/

Summary

4. Source code provisions in concluded UK Free Trade Agreements (FTAs) and digital economy agreements do not take sufficient account of the need of governments to introduce a range of measures that will regulate algorithmic and artificial intelligence (AI) systems, mitigate risks associated with the use of AI systems and ensure their developers and providers are held accountable for any harms that arise.
5. We recommend that the government:
   a) **Reviews its approach to source code provisions, carefully assessing whether there is sufficient concrete evidence to merit their continued inclusion in trade agreements.** In the event such provisions are retained, we recommend substantial modifications are made to leave room for the introduction of regulatory measures designed to address harms from AI and other digital technologies.
   b) **Looks for opportunities to renegotiate its existing treaty texts where source code provisions are narrowly drafted** (e.g. in the context of the upcoming CPTPP review).
   c) **Examines provisions on cryptography** to assess their potential impact on the regulation of AI and other digital technologies.
6. To date there has been minimal consultation on digital trade provisions beyond representatives from industry who had the opportunity to provide regular input through the (now disbanded) Trade Advisory Group on Telecoms and Technology. Other societal actors have had no similar mechanism to provide input and engage meaningfully.
7. We recommend the creation of a multi-stakeholder consultative group on digital trade which meets regularly to provide input to the Department for Business and Trade. In addition to business representatives, we recommend that the group includes experts from academia and research institutions, NGO and citizen's rights organizations, consumer groups, and trade unions. Similarly, the group should include government departments such as the Department for Science, Innovation and Technology, regulators such as those participating in the Digital Regulation Cooperation Forum, and bodies such as the British Standards Institute and the AI Safety Institute. In the interests of transparency and accountability, we recommend that the membership of such a group is made public and subjected to periodic review, and that a summary of discussions is published after each meeting. The consultations must ensure all stakeholders – not just the business community – are taken seriously and given an equal opportunity to contribute to shaping digital trade policy.

The importance of opening the 'black box' of AI systems

8. A growing number of everyday economic and societal interactions are mediated by algorithmic and AI systems (hereafter "AI"): the AI-as-a-service global market was valued at $40bn in 2022 and is growing rapidly. The AI supply chain is highly internationalised and UK firms are increasingly exporting and importing AI products and services. AI can enhance efficiency and productivity, but its widespread use can lead to individual and systemic risks. On an individual level, people experience a lack of transparency regarding how algorithms influence decisions made, feel disempowered in the face of highly technical, data-driven processes and suffer unfair treatment that may interfere with their legal rights but is nevertheless difficult to challenge or overturn. On a systemic level, AI can lead to harms linked, for example, to the growing concentration of market power in digital markets[2] or the erosion of trust in democratic institutions as a result of large-scale misinformation.[3]

9. High-profile examples of controversial, opaque and ultimately failed and unjust use of algorithmic decision-making in the UK public sector show the need for regulatory intervention and effective mechanisms for oversight and redress. We point to two: the systems used to determine A-level results during the pandemic[4] and the Home Office's withdrawal of a computer algorithm deployed to sort visa applications due to concerns it "entrenched racism and bias".[5] At the very least, higher levels of transparency are required, and the new government has signalled a commitment to deliver it through mandating compliance with the (so far voluntary) Algorithmic Transparency Recording Standard.

10. Growing calls for more effective government regulation and oversight of AI have intensified with the release of far more powerful foundation models and the rise generative AI. Legislators are moving to regulate AI: examples include the EU's AI Act and the UK Digital Markets, Competition and Consumers Act 2024. Regulatory interventions can be expected to rely heavily on AI standards, such as those developed at the national level, at the regional level by standards setting bodies such as CEN-CENELEC, and at the international level by organisations such as ISO/IEC.

---

[2] Competition & Markets Authority (2024), *CMA AI strategic update,* available at https://www.gov.uk/government/publications/cma-ai-strategic-update/cma-ai-strategic-update#introduction.

[3] Leslie, D. and Perini, A.M. (2024) 'Future Shock: Generative AI and the International AI Policy and Governance Crisis', Harvard Data Science Review, doi:10.1162/99608f92.88b4cc98.

[4] Kelly, A. (2021), 'A tale of two algorithms: The appeal and repeal of calculated grades systems in England and Ireland in 2020', British Educational Research Journal, 47, 725-741, https://doi.org/10.1002/berj.3705.

[5] Ungoed-Thomas, J. and Abdulahi, Y. (2024), 'Warnings AI tools used by government on UK public are "racist and biased"', The Guardian, available at https://www.theguardian.com/technology/article/2024/aug/25/register-aims-to-quash-fears-over-racist-and-biased-ai-tools-used-on-uk-public.

11. AI auditing will be an essential part of the AI governance regime, vital for ensuring that AI systems are legal, ethical and safe (for instance, meeting requirements for privacy, fairness/bias, explainability, and robustness). A private market of AI auditing firms is already emerging,[6] as firms and public sector buyers procuring and deploying AI systems want to know that they are fit-for-purpose. The previous government sought to establish the UK as a frontrunner in AI assurance, noting that the UK has 'the potential to excel' in this area.[7]

12. There is an ongoing debate among experts about exactly how much information external auditors, system buyers or their advisors require to conduct a robust assessment of an AI system. This debate is made more complex by the emergence and widescale use of foundation models and generative AI. Large-scale AI systems and their enormous capability and adaptability to perform a wide range of tasks create the need for new transparency and explainability frameworks and tools, as well as new evaluation and auditing methods. We do not have a full understanding of the capabilities and behaviours of large-scale AI systems, the transformer architecture that often underpins these systems, the massive data pools and the provenance of the data that the systems were trained on, the reasons why they produce certain outputs, or the range of uses and associated benefits and harms. **Without a better understanding of these factors – and others - it is not yet possible to determine what effective auditing and evaluation methods for foundation models and generative AI may look like or the level of disclosure that will be required**.[8]

13. It is clear, however, that the greater the level of access that an external auditor has to information about an AI system, the more thoroughly the system can be assessed.[9] External auditors may need access to an AI system's inner workings (e.g. weights, activations, gradients) – so called 'white-box' access - as well as access to training and deployment information (e.g. methodology, code, documentation, data, deployment details, finding from internal evaluations) – so called 'outside-the-box' access.[10] Moreover, access to the Application Programming Interfaces (APIs) of the

---

[6] Digital Regulation Cooperation Forum (2022), *Auditing algorithms: the existing landscape, role of regulators and future outlook*, available at https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook#the-role-for-audit-in-algorithmic-governance.

[7] Department of Business, Energy and Industrial Strategy and Department of Culture, Media and Sport (2021), *National AI Strategy*, available at https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf, 58.

[8] Tsamados, A., Aggarwal, N., Cowls, J. et al (2022), 'The ethics of algorithms: key problems and solutions', AI & Soc, 37, 215–230, https://doi.org/10.1007/s00146-021-01154-8.

[9] Koshiyama, A., Kazim, E., Treleaven, P. et al (2024), 'Towards algorithm auditing: managing legal, ethical and technological risks of AI, ML and associated algorithms', Royal Society Open Science, 11, 230859, https://doi.org/10.1098/rsos.230859; Casper, S., Ezell, C., Siegmann, C., et al (2024), 'Black-Box Access is Insufficient for Rigorous AI Audits', Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, https://doi.org/10.1145/3630106.3659037.

[10] Casper, Ezell, Siegmann *et al* (2024).

algorithmic or AI system is required for providing auditors with remote access and in order to carry out many standard tests that, for instance, control for biases.[11]

14. A major challenge for the conduct of thorough AI audits is that AI developers are wary of providing high levels of access to their models, due to concerns about possible disclosure of their trade secrets as well as the reputational repercussions of a bad assessment (e.g. if the results reveal lacklustre performance or poor risk-management processes). In addition, the current scope and level of access for foundation model evaluations is often directed and decided by the company or organisation that developed the model.[12]

15. As a result, external auditors typically rely on less rigorous black-box testing.[13] This can have counterproductive effects as poor quality audits can increase public or regulatory trust in systems on false grounds, preventing appropriate levels of external scrutiny. They also enable safety- or ethics-washing by developers who make AI systems that contribute to risks without sufficiently investing in methods to address them.[14]

16. In the absence of a legal requirement for firms to provide external auditors with the requisite access to their systems (which can be done in ways that protect developers from unauthorised disclosure of confidential information), it is hard to see how the government will be able to create an effective regime for AI regulation and oversight.[15] The recent experimentation with voluntary access commitments in the UK clearly shows the need for mandatory disclosure regimes.[16]

---

[11] Casper, Ezell, Siegmann *et al* (2024).

[12] Jones, E., Hardalupas, M., and Agnew, W. (2024) 'Under the radar? Examining the evaluation of foundation models', Ada Lovelace Institute, available at www.adalovelaceinstitute.org/report/under-the-radar/#how-should-regulators-and-policymakers-think-about-using-evaluations-25, 65.

[13] Casper, Ezell, Siegmann *et al* (2024).

[14] Casper, Ezell, Siegmann *et al* (2024).

[15] For the view on AI assurance as it stood in 2021, see Centre for Data Ethics and Innovation and Department for Science, Innovation & Technology, *The roadmap to an effective AI assurance ecosystem – extended version*, available at https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem-extended-version#roadmap-to-a-mature-ai-assurance-ecosystem. For an overview of the current challenges in the UK's AI assurance approach, see Davies, M., Strait, A., and Birtwistle, M. (2024) 'Safety first? Reimagining the role of the UK AI Safety Institute in a wider UK governance framework', Ada Lovelace Institute, available at https://www.adalovelaceinstitute.org/blog/safety-first/.

[16] Manancourt, V., Volpicelli, G., and Chatterjee, M. (2024) 'Rishi Sunak promised to make AI safe. Big Tech's not playing ball.' Politico, available at https://www.politico.eu/article/rishi-sunak-ai-testing-tech-ai-safety-institute/.

The ramifications of source code provisions in trade agreements for AI regulation and auditing

17. Source code provisions in trade agreements have important ramifications for algorithmic and AI auditing. In these provisions, governments commit not to require foreign companies to disclose source code, as a condition for the import, distribution, sale or use of that software, except in very limited circumstances. They were introduced into trade agreements in a bid to protect technology companies from so-called 'forced' disclosure requirements (a practice whereby some governments sought to acquire proprietary technology from abroad by requiring foreign companies to disclose proprietary information as a condition of doing business in their jurisdiction). Source code provisions feature in many recent digital trade chapters and digital economy agreements, including the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) to which the UK has recently acceded, the EU-UK Trade and Cooperation Agreement, as well as the UK's recent agreements with Australia, Singapore, and Ukraine.[17]

18. 'Source code' refers to instructions that a programmer supplies to the computer, typically to perform specific tasks. Access to the source code provides knowledge of how the software works "under the hood", which may often contain information crucial to a software's success.[18] From the regulator's perspective, accessing and studying source code is an important aspect of (some forms) of AI auditing[19] and hence valuable for enforcing AI regulations. It is particularly important for conducting *ex-ante* compliance checks to uncover undesirable practices before a product is deployed on the consumer market, or ahead of any modification of systems already in use. Without prior access to the source code, enforcement might primarily be *ex post facto*, where a breach only surfaces after harm is done.[20]

---

[17] The Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP), Article 14.17; 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 [2-21] OJ L-149 (EU-UK TCA), Article 207; Free Trade Agreement between the United Kingdom of Great Britain and Northern Ireland and Australia (UK-Australia FTA), Article 14.18; Digital Economy Agreement between the United Kingdom of Great Britain and Northern Ireland and the Republic of Singapore (UK-Singapore DEA), Article 8.61-K; and Digital Trade Agreement between the United Kingdom of Great Britain and Northern Ireland and Ukraine (UK-Ukraine DTA), Article 132-P.

[18] Mitchell, A.D., Let, D., and Tang, L. (2023) 'AI Regulation and the Protection of Source Code', International Journal of Law and Information Technology, 31, 283-301, https://doi.org/10.1093/ijlit/eaad026, 287.

[19] DRCF (2022) Auditing algorithms.

[20] Mitchell, Let and Tang (2023); Irion, K. (2022) 'Algorithms Off-limits?: If digital trade law restricts access to source code of software then accountability will suffer.' Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, https://doi.org/10.1145/3531146.3533212.

19. **There is thus growing tension between the core commitment in source code provisions in trade agreements, in which governments commit not to require firms to disclose source code, and the need for governments to mandate such disclosure as an integral part of an effective regime for governing algorithmic and AI systems and other emerging technologies**.

20. Trade agreements include 'exceptions', which allow governments to derogate from treaty commitments in specific circumstances. However, as we explain below, exceptions have, to date, been narrowly drafted. This limits the measures that governments can enact to require source code disclosure for the purposes of AI regulation without breaching their treaty obligations.

21. While there is mounting evidence that source code provisions impede effective AI governance, there is little compelling evidence that they provide additional effective protection to technology firms from so-called 'forced' disclosure – not least because the governments using this practice have not signed up to treaties including source code provisions. Moreover, source code provisions in trade agreements cannot normally be invoked by the owners, licensors or licensees of the source code and hence might not necessarily prove any more effective (if not less effective) than the protection provided in existing multilateral treaties (discussed below). In particular, intellectual property (IP) protection and the protection of trade secrets against unfair commercial use can offer more effective protection that private parties can rely on under domestic law. **While technology companies remain strong advocates of source code provisions, in an age of rapidly developing AI systems, it is no longer clear – from a public interest perspective – that the benefits of these provisions outweigh the costs.**

22. In light of these tensions, the UK and other governments have started to revise treaty text. Thus, for instance, the source code provision in the recent UK-Singapore DEA and the EU-Japan FTA are less restrictive than the CPTPP treaty which was negotiated almost ten years ago. However, problems persist even with more carefully drafted commitments. During the drafting of the EU AI Act, the EU had to reduce the scope of its regulatory ambitions to ensure coherence with commitments on source code provisions in its trade agreements.[21]

23. Since October 2023, the United States government has withdrawn its support for source code provisions in trade agreements (alongside provisions on data free flows, server locations and non-discrimination) citing concerns that they restrict the policy space required to regulate AI and other digital technologies effectively.[22] The New Zealand government has also stopped negotiating source code provisions in its

[21] Bertuzzi, L. (2023) 'How trade commitments narrowed EU rules to access AI's source codes', Euractiv, available at https://www.euractiv.com/section/artificial-intelligence/news/how-trade-commitments-narrowed-eu-rules-to-access-ais-source-codes/.

[22] Dupont, D. (2023) 'U.S. to end support for WTO e-commerce proposals, wants 'policy space' for digital trade rethink', Inside U.S. Trade, available at https://insidetrade.com/share/178191.

trade agreements (including the UK-New Zealand FTA) after the Waitangi Tribunal found the source code provision in the CPTPP did not adequately address Māori concerns about the risks of biased assumptions in algorithmic design and training data.[23] The UK and several other countries, including the EU, Australia, Singapore and Japan, have continued to include source code provisions. Although they have refined their treaty texts to try and address concerns this has not resolved all the challenges, as we explain below.

24. As the UK is only now developing a regime for AI governance and auditing, and the technology is evolving rapidly, **the UK government should carefully assess whether there is sufficient concrete evidence to merit the continued inclusion of these provisions in trade agreements**. Below we consider alternative mechanisms for providing UK technology companies with protection in overseas markets that do not restrict the scope for domestic AI regulation. **We also provide suggestions for how source code provisions could be redrafted. We also recommend that the government looks for opportunities to renegotiate its existing treaty texts where source code provisions are narrowly drafted (e.g. in the context of the upcoming CPTPP review)**.

25. Scholars also stress that there are inconsistencies between the source code provision in trade agreements and other legal regimes regarding the source code of software, including in the Berne Convention on the Protection of Literary and Artistic Works, the WIPO Copyright Treaty and the WTO Agreement on Trade-related Aspects of Intellectual Property Rights (TRIPS). For example, UK and EU copyright law recognise a right to reverse engineering as a legitimate means to discover the ideas and principles underlying the software .[24] These tools are essential to ensure interoperability and competition on downstream markets but they can be undermined by prohibitions on source code disclosure requirements. **In reviewing its approach to the source code provision, we recommend that the government undertakes a legal analysis of the consistency of the source code clause with existing laws protecting software source code and attendant exceptions for fair dealing.**

26. While we have not examined the cryptography provision in digital trade agreements, we note concerns that these provisions may also impede effective AI regulation. In these provisions, governments commit *inter alia* not to request firms to provide any proprietary information relating to cryptography or to use or incorporate

[23] Waitangi Tribunal Report (2023) *The Report on the Comprehensive and Progressive Agreement for Trans-Pacific Partnership*, available at https://forms.justice.govt.nz/search/Documents/WT/wt_DOC_195473606/Report%20on%20the%20CPTPP%20W.pdf.

[24] See UK Copyright, Designs and Patents Act 1988, Articles 50B and 50BA; Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (EU Software Directive) [2009] OJ L-111, Articles 5(3) and 6.

a particular cryptographic algorithm, except for national security reasons.[25] While encryption of private communications is crucial in safeguarding the privacy and freedom of speech of individuals, this commitment may impede the implementation of government cybersecurity measures specifying the types of cryptographic algorithms firms can use (for instance, measures requiring firms to upgrade their systems to interoperable quantum-resistant cryptography).[26] There is also the risk that AI developers encrypt their systems to evade scrutiny.[27] **For this reason, we recommend that the government reviews cryptography provisions in tandem with source code provisions to assess the ways in which they may invertedly impede the regulation of AI and other digital technologies.**

27. Below we elaborate on the challenges associated with the existing drafting of the source code provision and make specific suggestions as to how the provision could be modified.

Ambiguities regarding the scope of source code provisions

28. The source code provisions typically stipulate that Parties to the treaty will not require the transfer of or access to source code owned by a person of another Party. For instance, in the Ukraine-UK digital trade agreement, '*Neither Party shall require the transfer of, or access to, source code of software owned by a person of the other Party*' (Article 132-P.1). A footnote stipulates that '*For the purposes of this Article, a reference to "source code" includes an algorithm embedded in the source code, but does not include the expression of that algorithm in any other form, including in prose.*'

29. One concern with the drafting of source code provisions is that the key terms, including 'source code of software', are not defined in the treaty. In such cases, the Vienna Convention on the Law of Treaties (1969) governs interpretation and provides that the 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 (Article 31). However, we understand from computer science experts that there is no settled definition of source code in the expert community and hence no 'ordinary' or contextual meaning that can be relied upon. 'Code' is often used to refer to computer algorithms and machine learning models as these count as instructions that can be compiled and executed by a computer. As a

[25] See, for example, UK-Singapore DEA, Article 8.61-J.

[26] Rethink Trade (2024) *Limitations on Cryptography Rulemaking in Trade Agreements Could Generate Cybersecurity Risks*, available at https://rethinktrade.org/reports/memo-limitations-on-cryptography-rulemaking-in-trade-agreements-could-generate-cybersecurity-risks/.

[27] Jones, E., Kira, B., and Tavengerwei, R. (2024) 'Norm Entrepreneurship in Digital Trade: The Singapore-led Wave of Digital Trade Agreements', World Trade Review, 23, 208-241, doi:10.1017/S1474745624000089, 231.

result, the provision might reasonably be interpreted to include algorithms, training data, datasets, and other related components. This broadens the scope of source code provisions, creating ambiguity about what is truly covered, and poses even greater challenges for the regulation of algorithmic and AI systems.

30. A particular concern arises with application programming interfaces ('APIs'), as even a relatively narrow interpretation of the term 'source code' may reasonably include them. This would create a much bigger hurdle for AI regulation and auditing because APIs – if technically caught under the definition of 'source code' - include external-facing code meant for external parties to interact with the system.[28] Without access to APIs, it is hard to perform even the less intrusive (black-box) forms of AI auditing. Such an interpretation would not only affect the copyright reverse engineering rules discussed above, but also impact competition and competition law, since access to APIs and potentially other elements of the source code can be essential for ensuring interoperability and hence competitiveness of digital markets. **If source code provisions are retained, we recommend that they include a precise definition of 'source code' and explicitly state that APIs are excluded from scope.**

Narrow scope of exceptions

31. The scope and nature of exceptions that apply to the source code commitment vary across treaties as there have been substantial revisions to treaty texts in recent years. However, concerns remain that even the most recent drafting (e.g. in the UK-Singapore and UK-Ukraine digital economy agreement texts) may still unduly constrain the development of a robust regime for AI regulation and auditing. Moreover, the UK faces the challenge that some of its treaty obligations were drafted a decade ago, long before AI regulation and auditing became a policy priority (notably the CPTPP to which the UK recently acceded).

32. Three types of exception that apply to the source code commitment: treaty-wide exceptions, chapter-wide exceptions, and article-specific exceptions. The treaty-wide exceptions are typically modelled on GATS Article XIV and scholars have noted that these 'general exceptions' provide very limited scope for derogation from the source code commitment. To utilise general exceptions, one set of experts argues that "*AI regulators should judiciously limit the requirement of source code disclosure to exceptional situations where the AI system poses a high risk* to 'public morals' or 'public order' or where the protection of 'human, animal or plant life or health' is concerned and clearly specify their desired level of protection [emphasis added]."[29] Even if general exception clauses were interpreted to allow for more policy space, it is hard to see that this approach – whereby the inspection of algorithmic and AI

[28] Mitchell, Let and Tang (2023); Irion (2022).

[29] Mitchell, Let and Tang (2023).

systems is severely curtailed by obligations in trade agreements – is in the public interest. Rather than relying on general exceptions modelled on GATS Article XIV, we recommend that the UK government focuses on revising its approach to treaty commitments.

33. A second source of flexibility lies in chapter-wide exceptions. Agreements typically carve out government procurement from the digital trade chapter,[30] with the important implication that the source code commitments do not apply to government bodies procuring AI systems. **Given the sensitivity of AI systems used in government procurement, we recommend that the UK continues to carve out government procurement from the scope of digital trade chapters.** However, we also note that the procurement carve-out has limits that should be investigated further. We highlight two limits to the carve-out. First, because governments outsource important aspects of public provisioning to private providers, the exercise of some public functions may still be covered by the source code commitment. Second, the procurement carve-out may apply to interactions between the system's vendor and the procuring authority but, in the context of long digital supply chains, the relevant algorithm or AI may sit at a level where the carve-out does not apply.

34. The EU inserted an additional chapter-wide exception in its digital trade agreements in an article on the "right to regulate". This acknowledges the right to regulate in line with legitimate public policy objectives specific to digital trade notably privacy and data protection[31] **The UK could consider introducing a similar "right to regulate" article in its digital trade agreements explicitly specifying *inter alia* that addressing harms and risks associated with AI and other digital technologies is a legitimate public policy objective.** To make it more effective, such an article could introduce a binding language, such as "The Parties shall uphold and respect the right to regulate, including in addressing the harms and risks associated with AI and other digital technologies, as a legitimate public policy objective." This approach would not just reaffirm but oblige the Parties actively to protect the public interest.

35. Exceptions built into the source code provision itself (article-specific exceptions) are a vital source of regulatory flexibility, and they have grown more and more extensive as governments have come to recognise the challenges associated with the source code provision. The UK's most recent drafting is in UK-Ukraine Digital Trade Agreement, Article 132-P.2:

    "Nothing in this Article shall be construed to: (a) preclude a regulatory body or a judicial authority of a Party, or a designated conformity assessment body operating

[30] See, for example, EU-UK TCA, Article 207(20(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.'; UK-Australia DTA, Article 14.2-2(b): this Chapter does not apply to … 'government procurement'. The same text appears in CPTPP, Article 14.2-3(b).

[31] See EU-UK TCA, Article 198.

in a Party's territory, from requiring a person of the other Party to preserve and make available the source code of software in furtherance of an investigation, inspection, examination, enforcement action, or judicial proceedings; or (b) apply to a remedy imposed, enforced, or adopted by a regulatory body or a judicial authority of a Party, in accordance with a Party's law following an investigation, inspection, examination, enforcement action, or judicial proceeding."

The subsequent paragraph provides a commitment that, in implementing these measures, the Party will "prevent the unauthorised disclosure of source code of software'.

36. From a public interest perspective, this drafting reflects significant progress. In the CPTPP text, for instance, the exception only provides for governments to require the *modification* of source code to comply with laws and regulations and does not provide for the government to require *access* to source code, even though this is essential for AI regulation and auditing. US agreements (e.g. USMCA and Japan–US Agreement) were criticised for exceptions that only granted access to selected public bodies on a case-by-case basis, leaving no room for *ex-ante* regulation and oversight such as that introduced in the EU's AI Act.[32] As reflected in the UK-Ukraine text above, more recent texts are drafted to provide for both *ex ante* regulation and *ex post* remedial actions. They also expand the list of authorities that can require firms to preserve and make available source code, including conformity assessment bodies. Recent EU texts (including the EU-UK TCA and the EU-Japan agreement) go further in specifically excluding certain types of government action from scope, including requirements related to competition law, while the Singapore–UK agreement specifically exempts measures required for monitoring compliance with codes of conduct and standards.[33]

37. Nevertheless, in this paragraph, we highlight four areas of concern where the exceptions, as drafted in recent texts, may still create barriers to the creation of an effective regulatory framework that guarantees appropriate levels of accountability and transparency for those affected by AI and algorithmic systems and their representatives. The following is based on the UK-Ukraine text:

(a) The agreement does not define 'designated conformity assessment bodies': The lack of a definition creates uncertainty, particularly in areas where the UK government is hoping to rely on private actors for AI auditing and assessment as it is not clear whether they would fall into the category of 'designated conformity assessment body'.[34] **The source code provision could be reworded in functional terms, rather than operator-based terms. Through such a re-drafting, the provision**

[32] Jones, Kira, and Tavengerwei (2024).

[33] Jones, Kira, and Tavengerwei (2024).

[34] See footnotes in paragraph 7 above.

could reference conformity assessments carried out by or on behalf of a regulatory body or a judicial authority.

(b) The agreement does not cover regulatory measures requiring AI firms to cooperate (and hence disclose information) in the context of pre-market, *ex ante* assessment of algorithmic and AI systems that are non-judicial or not undertaken by a 'regulatory body' or 'conformity assessment body'. There are various groups of non-judicial stakeholders that may reasonably seek to assess digital systems before they enter the UK market or before they are implemented in a particular setting. Worker representatives or consumer rights organizations would be two examples that currently face barriers to transparency in advance of the use of a system. Gaps also exist in the public sector. Government/local offices or authorities that are not either (1) regulatory bodies or (2) a conformity assessment body or (3) implementing a decision of such a body would also fall outside of the permitted exception cited above. **Careful consideration is needed regarding which government/local authorities or other non-state actors might need to be the beneficiaries of transparency and explainability regarding AI and algorithmic systems in the context of the UK's future regulatory regime. A wording for the source code provisions must then be found that ensures that their activities can proceed without international treaty provisions creating additional hurdles.**

(c) The agreement does not cover regulatory measures requiring AI firms to cooperate (and hence disclose information) in the context of non-judicial resolution of disputes, for example through arbitration, mediation or settlement processes. Information about AI and algorithmic systems can be central to a range of disputes, not all of which are resolved in a judicial forum. One example of this is online dispute systems between users and platforms where the user requires transparency and an explanation of AI-based decisions to moderate, recommend or remove content. Particularly where there is already an information asymmetry between the parties to the dispute (for example, consumer-business or worker-employer), the source code provisions may present a barrier to seeking appropriate disclosure during the non-judicial resolution of such disputes. This barrier would only serve to increase the information asymmetry and thereby damage the capacity of the relevant tribunal/forum to achieve justice or resolve the dispute fairly. As above, **a wording for source code provisions should be crafted that enables the fair resolution of disputes in non-judicial fora.**

(d) The agreement does not explicitly reflect fair dealing and other public interest exceptions set out in current UK IP (including copyright) and trade secrets law. This lack of recognition creates an inconsistency between the source code provisions and IP law relevant for AI and other uses of source code. **Any future approach to the source code provisions should ensure alignment with the protection and limits of IP law.**

38. Section 4 of the UK-Ukraine provisions reads: "This Article shall not apply to the voluntary transfer of, or granting of access to, source code of software by a person of the other Party: (a) on a commercial basis, such as in the context of a freely negotiated contract; or (b) under open source licences, such as in the context of

open source coding." This drafting may also prove to be narrow in the context of public procurement particularly. A term in a procurement contract may be imposed by the public buyer, rather than being susceptible to negotiation during the tender procedure. Therefore, it may not be considered to be 'freely negotiated'. An alternative drafting (as in Article 12.11 of EU-New Zealand FTA) refers to transfer/grants of access 'in the context of a public procurement transaction or a freely negotiated contract'. An outstanding challenge, even on this wider drafting, would be that some procurement-related requirements may emerge years after the relevant procurement contract (or "transaction") is concluded. **Language that applies the exemption to such requirements, which may be unforeseen in the original contract, and responds to the concerns raised at [29] above, should be sought.**

Source code provisions: two options for the future

(1) Remove the source code provision and modify existing "trade secrets" articles

39. Given the challenges associated with the source code provision, one option is to simply discontinue, akin to the current approach of the U.S. and New Zealand, the inclusion of source code provisions in digital trade agreements.

40. Existing multilateral commitments under the WTO TRIPS Agreement (Article 39) and the Paris Convention for the Protection of Industrial Property (Article 10bis) already provide for a solid basis of protection of confidential information of AI and other technology developers against disclosure and against unfair commercial uses. Integrating compliance with these multilateral treaties into digital trade agreements gives them additional "teeth". Indeed, trade secrets articles are already found in IP chapters of UK FTAs (e.g. Article 17.63 of the UK-New Zealand FTA). These articles could be modified to include only access to, disclosure and use of source code that is 'contrary to honest commercial practices' (as often further defined in the FTA and Article 10bis of the Paris Convention). In that way, IP protection for source code can be tailored to include relevant limits and exceptions, such as the legitimate means to achieve interoperability, the need for AI transparency and explainability, and other legitimate activities of users and competitors. In particular, provisions should be drafted explicitly to exclude regulatory and other legitimate activities designed to address the harms and risks arising from AI and other emerging technologies. This aligns with the general principle that IP rights as private rights do not interfere with the State's ability to regulate.[35]

41. Removing the source code article and modifying trade secrets articles in the manner we propose above would secure a significantly wider policy space to regulate key emerging technologies such as AI while retaining robust protection of businesses against unfair practices. It would of course hinge on the effective availability of such

35 See the Preamble and Articles 8, 13-14, 16, 28, 30, 39 of the WTO TRIPS Agreement and the WTO dispute settlement case law, such as the Appellate Body Report in Australia – Plain Packaging, 2020.

protection under the domestic law of the trading partner and hence raise the well-known problems and pitfalls of effective IP enforcement abroad. However, this approach has the advantage of being consistent with existing legal regimes under which private businesses are responsible for enforcing their IP rights and trade secret protection both under applicable domestic laws at home and abroad.

(2) Modify current source code provisions

42. If the UK government considers it is essential to continue with the inclusion of source code provisions we recommend further revisions to its approach, building on the progress made to date. In line with the discussion above, such modifications should, *inter alia*:

- expressly and narrowly define 'source code of software', in so doing, clarify that APIs are out of scope;
- reaffirm that existing fair dealing exceptions and public interest exceptions in domestic IP law and trade secrets law continue to apply;
- adopt a functional approach to the exemption for conformity assessments (and other audit or evaluation measures as needed);
- allow for a wider range of methods for the audit and evaluation of AI and algorithmic systems, particularly those involving non-state actors (including firms, consumers, workers and researchers);
- ensure that the source code provisions allow for the introduction of government measures requiring the disclosure of information about algorithmic and AI systems to protect the rights and interests of private individuals (firms, workers, and consumers), or their representatives on their behalf, in accountability settings beyond the regulatory and judicial context that is currently mentioned.

43. To avert the need to continually update the exceptions to the source code provision to keep pace with government regulation of fast-moving technologies, **a broader regulatory carve out could be incorporated, taking inspiration from the precedent the EU has set in carving out regulatory measures for personal data protection in its digital trade agreements**. For instance, the EU-New Zealand FTA, Article 12.5(2), states that '*Each Party may adopt or maintain measures it deems appropriate to ensure the protection of personal data and privacy, including through the adoption and application of rules for the cross-border transfer of personal data. Nothing in this Agreement shall affect the protection of personal data and privacy afforded by the Parties' respective measures'*. This would ensure that the government has the scope it needs to effectively regulate algorithmic and AI systems, and other digital technologies.

Conclusion

44. When adopting policy in this area of rapid evolution, governments must balance the potential benefits of the source code provisions with the adverse impact that source code provisions have, or may have in the future, on the government's capacity to regulate algorithmic and AI systems and other digital technologies. It is far from clear that the purported benefits of source code provisions outweigh the serious adverse effects upon a government's scope for regulatory manoeuvre. Given the significant potential for source code provisions to apply in unintended ways to new technologies, we primarily suggest that government considers the discontinuation of source code provisions. If it is felt that retaining these provisions is essential, we suggest significant modifications that guarantee sufficient policy space to regulate emerging technologies of today and tomorrow in the interest of accountability.

Authors

Emily Jones is an Associate Professor in Public Policy at the Blavatnik School of Government and a Fellow of University College, University of Oxford. She is a co-founder and co-director of the Trade and Public Policy Network and recently served as a special adviser to the International Trade Select Committee in the UK Parliament.

Philippa Collins is a Senior Lecturer in Law and Co-Director of the Centre for Law at Work at the University of Bristol.

Henning Grosse Ruse-Khan is a Professor of Law and Co-Director of the Centre for Intellectual Property and Information Law, University of Cambridge. He is also a Fellow at King's College, Cambridge.

Albert Sanchez-Graells is a Professor of Economic Law at the University of Bristol. He is also a member of the Cabinet Office Open Contracting Advisory Group and former member of the European Commission Stakeholder Expert Group on Public Procurement.

Kristina Irion is an Associate Professor at the Institute for Information Law, University of Amsterdam, and a Guest Lecturer at the University of Lucerne.

Cosmina Dorobantu is the Co-Director of Public Policy Programme and Policy Fellow at The Alan Turing Institute and Turing OII Fellow at the Oxford Internet Institute, University of Oxford.

Burcu Kilic is a Senior Fellow at the Centre for International Governance Innovation and a tech and human rights Fellow at the Carr Center for Human Rights Policy at the Harvard Kennedy School.

Daria Onitiu is a Post-doctoral Researcher at the Trustworthiness Auditing for AI project and Programme on the Governance of Emerging Technologies (GET) at The Oxford Internet Institute, University of Oxford.