AI Whitepaper – UK Finance response

Date: 28 June 2023

Sent to: evidence@officeforai.gov.uk

UK Finance is the collective voice for the banking and finance industry.

Representing more than 300 firms across the industry, we act to enhance competitiveness, support customers and facilitate innovation.

Key messages from our response include:

- We support the sectoral, risk-based approach, based around regulatory guidance. Such an approach provides flexibility and can account for the peculiarities of each sector, use case and the existing applicable rules more readily than horizontal AI law. It is also more able to adapt to technological developments than primary legislation.
- The UK government and the central function can play a key role in convening and coordinating regulators, and facilitating the sharing of information. The design of this central function should account for the following:
    - Where guidance covers cross-sectoral issues, there should be coordination to promote coherence and avoid undue duplication, fragmentation or contradictions, with joint guidance pursued where feasible. This will help maintain a level playing field between sectors and promote innovation. Although the policy focus is rightly on outcomes, where there is a need for technical, process- or governance-focused guidance, there may be more cross-sectoral elements, requiring more coordination.
    - The government and a central function will not be able to direct or override regulators (absent a new statute). Care is needed to preserve regulatory independence and any statutory 'hook' like a duty to have due regard to a set of principles needs to be considered carefully, in light of each relevant regulator's statutory duties and role.
- It is important to clarify in the final policy papers that regulators can rely on their existing, technology-neutral guidance and rules, with tactical AI-specific supplements when required. They should not be expected to produce a full 'AI overlay' when existing regulation addresses risks adequately. AI frameworks should consider existing regulation and carefully evaluate gaps in order to avoid regulatory duplication, while ensuring high risk use cases are not overlooked. Avoiding unnecessary new layers of guidance will help reduce complexity and promote innovation.
- The UK approach to AI needs to consider not only risks posed by misuse or error on the part of legitimate firms but also the risks posed by bad actors using AI for fraud or other malicious purposes. Similarly, the potential for AI to contribute to other policy priorities needs to be considered, for example its use in fraud prevention.
- There is a mix of principles-based and prescriptive approaches emerging globally. This will make international interoperability challenging but government should seek to promote it and find opportunities to drive alignment (or at least compatibility) when possible, diverging from international norms only when there is a good reason to do so.
- We strongly support the development of sandboxes as a tool for delivering business certainty and revealing any areas of tension between regulatory expectations.
- Generative AI – where made publicly available online – merits particular attention from policy makers and regulators, given the new risks and challenges it poses (alongside clear opportunities for businesses and consumers). This is a potential gap in existing regulation that should be reviewed as a priority.
- Monitoring of supply chain issues will be needed, for example to ensure that firms deploying AI (and holding most regulatory obligations) are able to access the information

they need to do so with confidence while respecting vendors' intellectual property. Cross-sectoral AI assurance tools and documentation should be developed, building on the existing work of the Centre for Data Ethics and Innovation (CDEI).

In the context of discussing the issues raised in the Whitepaper, we also put forward a number of suggestions that may be most appropriately considered by regulatory authorities in due course, rather than being for the government to address in its overarching approach. We nonetheless include these, as being relevant to the thinking about how the regime will operate in practice.

Please find our detailed responses to the consultation questions annexed.

If you have any questions relating to this response, please contact Walter McCahon, Principal, Privacy and Data Ethics at walter.mccahon@ukfinance.org.uk.

Walter McCahon
Principal, Privacy and Data Ethics

Annex – UK Finance responses to AI Whitepaper consultation questions

The revised cross-sectoral AI principles

Question 1 – Do you agree that requiring organisations to make it clear when they are using AI would improve transparency?

1. UK Finance supports the *spirit* of this idea, as we see transparency as vital to maintaining public trust in the technology.
2. As an initial comment, the Whitepaper suggests that proportionality will be applied: "An appropriate degree of transparency and explainability should be proportionate to the risk(s) presented by an AI system." We fully endorse this approach but note that this may not be straight forward in practice. In particular, an AI vendor may (reasonably) take a view that its model does not present significant risks, but the way it is deployed by the client firm might. In this situation there could be a mismatch between the views of the vendor and the client, potentially leaving the client without the necessary insights into the model. See also our comments under Questions L1 to L3.
3. However, we are unsure what government is in fact proposing or the intent behind the proposal, noting a number of subtleties and challenges to work through. In particular:
    a. How does this fit with the overall model of sectoral regulators taking the lead and setting expectations for use cases within their domain? Discussion of "requiring" firms to do something implies some kind of horizontal rule applied to firms directly by government, such as a new statutory requirement. This would contradict the overall framework in the Whitepaper, if indeed intended, seemingly cutting across the transparency principle.
    b. We are similarly unsure of the policy rationale behind this question. The policy goal will inform the approach taken, with different approaches warranted depending on whether the goal is to build customer trust, ensure regulatory compliance through communications to authorities or indeed if there is some other objective.
    c. Building on this point, it is unclear from the Whitepaper text to whom firms would need to make their AI use clear, or in relation to what use cases. We can understand the goal of ensuring that consumers do not mistakenly believe they are interacting with a human in the context of a chatbot for example. But many AI use cases are used in back-office administration; in our view, consumers would not want to be told about details such as business optimisation software or the basis on which banks make macro level capital management decisions.
    d. Clearly, regulators will need sight of certain back-office AI uses but the relevance of different applications will likely vary by sector.
    e. Alternatively, we note that the intention behind this provision in the Whitepaper might be for firms to issue some kind of public disclosure of certain back-office uses of AI in order to raise public awareness. This could potentially take the form of a public statement, or an explanation in the company accounts or privacy notice, or – where relevant – some kind of in-app notification. If this were intended, it would need careful consideration and would need to allow flexibility for different use cases, contexts and sectors, while avoiding the risk of 'notification fatigue' among users.
    f. The right approach in our view surely depends on how the AI is deployed; as the implications would vary, sectoral regulators seem best placed to decide where such disclosure rules are warranted.
    g. Even in the context of directly consumer facing applications like chatbots, relationship management tools or loan decisions, we would ask whether the key issue here is in fact whether the consumer is engaging with 'AI', or whether the consumer is engaging with an 'automated system'. Is this about being transparent that the consumer is not interacting with a human being, or about AI specifically? We recognise that AI will probably be subject to increased public concern compared to more traditional automated systems, likely being perceived as posing greater

risks of potential customer harm or manipulation, but the policy rationale needs to be clear. We also highlight that guidance from the Information Commissioner's Office (ICO) will set minimum expectations for (significant) automated decision-making information provision to data subjects under the draft Data Protection and Digital Information (DPDI) Bill already. As such, this proposed requirement may be redundant in relation to consumer-facing use cases.

4. In light of the above, ***we do not agree with the creation of a new statutory transparency obligation***, if this is in fact being proposed. We agree that transparency is important but consider that Article 22 of the UK's General Data Protection Regulation (GDPR), supplemented by ICO guidance, plus rules and guidance from sectoral regulators already provide a suitable solution that allows for flexibility to account for differences in use case and audience.
5. Designing transparency, education and notification mechanisms that build trust requires a careful and context-sensitive design approach.

Question 2 – Are there other measures we could require of organisations to improve AI transparency?

6. As a general matter, we support a sectoral approach to AI frameworks that is risk- and principles-based. Per our comments above, the government's intent behind the transparency question is unclear and implies an interest in passing a new statutory requirement, which does not fit with the overall sectoral, guidance-based framework.
7. That said, we note that an area for regulators to consider will be the extent of specific AI reporting they want to see from firms they regulate (for those regulators possessing powers to require such regulatory reporting).
8. Regulatory reporting may assist where sectoral uses are not well divided and there is risk of overlapping or contradictory guidance from different regulators. We recognise that AI reporting could allow regulators to cooperate more effectively by understanding when AI use cases in their sector have shifted into the purview of a different regulator. We also recognise that the nuances of developing comparative baselines and use cases to accurately audit will be a difficult delivery for authorities.
9. Intelligence on current AI uses in their domain, obtained via regulatory reporting by firms, would help support risk-based regulation, which we support.

<u>Supply chain considerations</u>

10. A further important question remains about who needs to be transparent and how transparent they need to be. Must both developers and deployers of AI provide transparency and explainability information? As touched on elsewhere in our response, there is an important question around whether firms deploying AI can readily access the information needed from AI vendors (or indeed former vendors) for due diligence purposes, while managing legitimate vendor concerns about intellectual property.
11. However, existing regulators may not have powers to directly impose transparency obligations on vendors selling into their sectors. Work by the CDEI on the AI assurance ecosystem may be able to play a valuable role, here.[1]
12. See also our comments under Questions L1 to L3.

[1] See for example: https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem and https://www.gov.uk/guidance/cdei-portfolio-of-ai-assurance-techniques

Question 3 – Do you agree that current routes to contest or get redress for AI-related harms are adequate?

13. We believe that current regulation, such as GDPR's provision of a right to human review for automated decision-making (ADM), goes a long way to providing redress over many of the key issues relating to AI. Further layers of redress will be sector dependent. While the financial services sector has the detailed Dispute Resolution rules of the Financial Conduct Authority (FCA) and the Financial Ombudsman Service (FOS), other sectors may need strengthening in due course. (Though we note that the FOS would likely need upskilling and resourcing in order to effectively deal with AI issues, like other authorities).
14. We note that it will be necessary to develop best-practice for ADM redress and understand that the ICO will produce guidance on the updated ADM rights under UK GDPR after the passage of the DPDI Bill.
15. That said, we recognise that identifying 'AI-related harms' may not always be straight forward, where an AI tool is used to do analysis or produce outputs at a macro level, which are subsequently used lower down as an input into decisions that impact customers directly (potentially with other intermediate levels of analysis). This will require monitoring.

Question 4 – How could current routes to contest or seek redress for AI-related harms be improved, if at all?

16. We would revert to our above answer in Question 3 and note that further exploration of liability management issues is necessary to resolve this point, as liability allocation is complex. See also our comments below on liability and supply chain issues under Questions L1 to L3.
17. We note that there are some complexities over how certain UK GDPR rights and obligations apply / can be applied in some AI scenarios, such as where personal data can be inferred from a model. We note, however, that changes via the DPDI Bill will help provide clarity on the definition of 'personal data'. The ICO also has an active workstream on anonymisation, which can assist, as can the growing domain of privacy enhancing technologies, which offer the potential to significantly mitigate privacy risks in relation to models.[2]
18. We also note the development in the EU of a directive on the 'Liability rules for Artificial Intelligence'. Allowing the development of caselaw may resolve liability challenges. Nonetheless, we suggest close monitoring of the effectiveness of these EU measures to determine whether a similar approach might be warranted in the UK in due course.

Question 5 – Do you agree that, when implemented effectively, the revised cross-sectoral principles will cover the risks posed by AI technologies?

Overarching approach and comments

19. We support the view that an AI governance framework led by existing regulatory authorities and based on existing laws will best cover the risks posed by AI technologies and provide clarity to business, giving companies sufficient ability to experiment and innovate and leveraging existing regulators' expertise.
20. We also support the principles put forth in the Whitepaper, though we do note there is some uncertainty as to how these should best be applied by government without undermining regulatory independence, as set out under Questions 9 and 14.
21. Nonetheless, we prefer this model, based on context-specific cooperation, to creating a cross-sectoral AI regulator, agreeing with the reasons in section 3.2.3 of the Whitepaper.

[2] See in particular: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/privacy-enhancing-technologies/

The principles

22. In relation to the specific principles put forward we note that we are limited in the critiques we can make, as the detail of how they will be applied by sectoral regulators remains unknown for now. Nonetheless, we do have observations and recommendations, though these may be most relevant at later stages, when authorities are preparing their approaches and any guidance:
    a. We note that privacy is not included as a principle, though violation of the right to privacy is mentioned as a risk in places. We presume that this is because data protection legislation is taken to provide adequate protection already under the ICO, as the data protection regulator. We do not disagree per se, but we note that regulators will at times need to consider trade-offs between privacy and other principles such as fairness, which they should be encouraged to do transparently, acknowledging where balancing has been necessary. Similarly, privacy trade-offs should be considered transparently – and in discussion with the ICO – in situations where regulators may wish to see firms deploy AI to achieve other policy goals, such as identifying and protecting vulnerable individuals. The ICO recognises such trade-offs to an extent in its existing AI guidance, but we are not aware of such trade-offs being deeply considered elsewhere.
    b. 'Human in the loop' (or on the loop, etc) may be a helpful additional example under the governance principle in annex one, though it would not be appropriate in all cases.
    c. In relation to 'fairness' in annex one bullet four states: "Where a decision involving use of an AI system has a legal or similarly significant effect on an individual, regulators will need to consider the suitability of requiring AI system operators to provide an appropriate justification for that decision to affected parties." Insofar as personal data is involved, this duplicates the safeguards in UK GDPR (post-DPDI Bill). It would make sense for sectoral authorities to help ICO *finesse* relevant guidance when there are sector-specific considerations but the wording in annex one seems to allow for a situation where a sectoral regulator could contradict the ICO in relation to a GDPR requirement, which is surely inappropriate.
    d. In relation to the 'transparency' principle and its application in practice:
        i. 'Adequate transparency' should be assessed on the basis of the materiality of the model, the associated inherent risks related to the design, theory and logic of the model, the use case and the wider context.
        ii. Transparency is not an end in and of itself; it is a means of enabling accountability, empowering users, and building trust and confidence. In designing transparency guidance, regulators will need to consider their precise objectives and how best to meet these in a given context.
        iii. Different stakeholders will require different forms of transparency under different circumstances. Information given to stakeholders will need to be actionable, arrive at a relevant time be comprehensible and avoid extraneous details that can be distracting or confusing.
        iv. Regulators will also need to carefully balance the desire for transparency with other priorities, which might at times be in tension, for example speed, safety, security and privacy. Importantly, some forms of transparency can carry risks to user privacy and to firms' intellectual property but may provide little actual benefit in terms of enabling accountability and building trust. (See also our comments below under Question L1).

Application of the principles

23. In relation to the application of the principles, we also note that:
    a. In light of discussion in annex one of the Whitepaper and paragraph 63, it is currently unclear whether the government expects:
        i. each regulator to produce an 'AI guidance book' with AI-specific guidance on each of the principles, or

ii. each regulator to *reflect on* the AI principles and ensure that their rules and guidance adequately cover AI risks.

b. ***In our view, it should be made clear that (ii) is an acceptable approach***. This would enable authorities to rely on their generic (technology neutral) rules and guidance when these are sufficient, leveraging existing laws to avoid duplication and confusion. For example, the FCA and Prudential Regulation Authority already have extensive guidance on model risk management, fairness and protecting vulnerable customers. Similarly, the ICO's approach to fairness – centred around avoiding unjustified adverse impacts – is already highly relevant to AI. In financial services, the FCA already has in place the 'consumer duty', which broadly speaking requires firms to act to deliver good outcomes for retail customers and is supported by extensive guidance.

c. It should not be *necessary* for each regulator to layer on top an additional AI guidance book when the existing tech-neutral guidance and rules adequately address key risks. Similarly, when cross-sectoral guidance covers an AI risk effectively, it should not be expected that individual regulators apply their own layer. This would contradict the outcomes-focused approach, and risks duplication and unnecessary complexity. Of course, where there are *specific gaps or points of uncertainty* relating to certain types of model, system or use case, it would be logical to fill these.

24. Furthermore, the final principles should recognise that there can be trade-offs between the principles, which regulators will need to work through when setting guidance.

<u>Accounting for different users and customers</u>

25. Regulators should have scope to set out different expectations in relation to different classes of customer or other stakeholder. For example, more sophisticated customers – such as certain types of business customers – may have lower or different information needs, or different needs and expectations in relation to 'fair treatment', than retail customers. Where regulators are able to make relevant distinctions between types of stakeholder, they should be able to nuance their expectations accordingly.
26. We presume that regulators would have scope to make such distinctions where appropriate and final government papers should acknowledge this.

<u>Cross-sectoral guidance</u>

27. Where a concern is cross-cutting, an approach under which regulators work together to develop guidance that can be applied on cross-sectoral basis would be desirable. This would allow a collective view to be established, removing the possibility of duplicative effort and conflicting expectations. We anticipate that this may be most likely in the case of technical, governance- or process-focused guidance, although the policy focus is rightly on outcomes.
28. Where priorities are necessarily different, collaboration would help surface and resolve tensions.
29. We recognise however that in particular cases one regulator may be 'first among equals', due to its specific legislative mandate, its expertise or its resources. For instance, in the case of fairness – at least in relation to discrimination and other human rights issues – the Equality and Human Rights Commission (EHRC) could take the lead in developing guidance for scrutiny by (in a Financial Services context) the FCA, to identify potential nuances for this sector that need to be accounted for. See also our comments under Question 12.1 and Question 14.
30. Following on from this, the potential for one authority to act as 'first port of call' on specific cross-sectoral AI regulatory issues should also be explored.
31. We note of course that any arrangements under which a given authority takes the lead on a given issue would need to respect and conform to the powers, duties and scope of each authority under its establishing legislation. MOUs or other, more formal tools, can assist in building a coherent framework for managing overlapping domains efficiently.

Question 6 – What, if anything, is missing from the revised principles?

Overall approach

32. We support the Secretary of State's statement that the UK government does not intend to introduce horizontal AI legislation imposing duplicative rules on firms, but instead intends to rely on a principles-based framework. Strong messaging from government over time that this remains the case will be important to provide assurance to companies developing and using AI in the UK.
33. As we have stated previously, the financial services sector (including payments) is highly scrutinised and regulated, both at national and supranational levels. We believe that much of the existing legislation can be applied or adapted to apply to AI, and that this approach would be the most efficient route to effective outcomes in many cases, supporting competition and innovation by avoiding the imposition of undue regulatory burden for businesses investing and innovating.

Definition of AI

34. As we have outlined elsewhere in our response, we think that regulators should be able to rely on technology-neutral guidance when this is appropriate.
35. However, where guidance is to be focused on a given technology, there needs to be a clear definition in order for firms to know which systems to apply the guidance to, or at least prioritise. It may be appropriate at times for guidance to be focused on specific types of AI model or system (for example, neural networks or LLMs), in which case a clear definition of the relevant techniques will be needed.
36. Where guidance is in relation to AI in general, a clearer definition than what is laid out in 3.2.1 of the Whitepaper will be needed. Ideally this definition should be common across regulators in order to avoid distortions, an unlevel playing field or confusion. 3.2.1 goes in a positive direction but would need to be made more concrete for use in regulation or regulatory guidance. Notably, a regulatory definition may need to incorporate the notion of a 'system' or a 'model'.[3]

International interoperability

37. We believe that building a concept of equivalence or mutual recognition of other countries' AI regulatory regimes would be welcome. Such mutual cooperation will allow AI providers to work across borders with confidence to deploy their skills in tacking global issues such as healthcare improvement or climate change. A network of barely compatible regulatory regimes will not support the solving of globally important problems.[4]
38. (See also our comments on privacy under Question 5).

[3] See for example:

OECD definition of an 'AI system' at https://oecd.ai/en/ai-principles

Definitions of 'model' from the United States Federal Reserve at https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm and the UK Prudential Regulation Authority at https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/supervisory-statement/2023/ss123.pdf

[4] We note that interoperability was a main feature in the recent G7 whitepaper: https://www.csis.org/analysis/advancing-cooperative-ai-governance-2023-g7-summit

A statutory duty to regard

Question 7 – Do you agree that introducing a statutory duty on regulators to have due regard to the principles would clarify and strengthen regulators' mandates to implement our principles while retaining a flexible approach to implementation?

39. In principle this may be a suitable approach.
40. We agree that this would likely strengthen the principles' potency and encourage their continuing adoption. We note that regulators are bound by their statutory mandates; government cannot instruct them to act in particular ways outside of the arrangements laid down by statute.
41. However, we note that there are a number of important points to think through regarding how this would work in practice before we can endorse such a legislative measure. In particular, how would this statutory duty be intended to interact with regulators' existing statutory roles and objectives?
42. We presume that these would remain the primary objectives, which would 'trump' the AI 'due regard' obligation. Conceivably, the principles could be to 'colour or inform' how regulators interpret and pursue their statutory goals, or they could be similar to the 'secondary objectives' of some authorities.
43. An example could be a model like new section 120B of the Data Protection Act, to be created by the Data Protection and Digital Information (No 2) Bill. This requires the Information Commission to "have regard to" certain goals when "carrying out functions under the data protection legislation", when the Commissioner considers them "to be relevant in the circumstances".
44. An alternative model could be the duty on public authorities to "have due regard" to a series of anti-discrimination objectives under section 149 of the Equality Act. However, we note that this provision is quite detailed and includes a range of exemptions in Schedule 18; any comparable AI requirement would need equally careful design.
45. We also note that adding five new objectives to regulators' objectives – albeit secondary objectives – may risk diluting or confusing their mandates.
46. **We recommend considering in detail how such a statutory duty – or other legislative obligation – would interact with regulators' existing statutory mandates and whether there might be any tensions**. This could be complex, given the range of different authorities that would presumably be in scope so nuance and exemptions might be required, perhaps in a manner similar to the above Equality Act obligation.
47. If it appears that a non-statutory framework is achieving its objectives – with government simply convening authorities and publishing the cross-sectoral principles (and possibly other materials) as *informational guidance* – **a statutory approach may not be required**.
48. Government will need to keep under review whether cross-sectoral cooperation and consistency of approaches to non-sector specific issues is satisfactory. This will be an important component in the success of the proposed regime, as discussed further elsewhere in our response (e.g., Question 5 and Question 9).

Question 8 – Is there an alternative statutory intervention that would be more effective?

49. Building on our response to Question 7, statutory intervention should be considered carefully. The implications and interactions with other laws should be considered carefully, with due time given to see how well the non-statutory approach is working.
50. A further legislative tool worth considering further could be a 'handbrake' power, enabling government to place a temporary moratorium on the deployment of a certain AI use case, where this meets certain extreme risk criteria. The potential for such a power would require further careful policy analysis, however. This would need to cover for example the need to incorporate clear criteria, and appropriate obligations on government to consult the public and regulatory authorities, and to consider the potential up-sides of the use case.

Consideration would also be needed of appropriately clear timing provisions around the imposed moratorium and a process for consulting regulators and firms to resolve concerns.

51. Lastly, though it might not require a statutory intervention, whistle-blower systems could be set up by regulators so that staff at developers and deployers of AI can raise concerns effectively.

New central functions to support the framework

Question 9 – Do you agree that the functions outlined in Box 3.1 would benefit our AI regulation framework if delivered centrally?

52. Yes, overall, we agree with the proposed functions and their likely benefits.
53. We particularly favour measures to minimise duplication of rules and guidance, and to avoid any conflicts between the expectations of regulators, where their domains overlap. Having structures in place to promptly surface areas of apparent tension and then resolve these will be a key requirement for the success of the regime. Similarly, it will be key to actively monitor for high-risk use cases that risk 'falling between the cracks' outside of sectoral authorities' domains (though presumably still subject to ICO and EHRC rules, insofar as individuals and their personal data are concerned).
54. We note that the exact role of the central function – and how it will interact with regulators – needs to be fleshed out carefully. In particular, the vital principle of regulatory independence needs to be maintained. As such, presumably the central function can play a valuable role in convening regulators, hosting discussions, identifying points of tension and facilitating the sharing of information, but cannot *override or direct* sectoral authorities' decisions. (This would presumably require primary legislation if it were to become the policy intent). See also Question 14.

Question 10 – What, if anything, is missing from the central functions?

Additional activities

55. International interoperability
    a. We support the goal of having the central authority encourage international interoperability. The EU and the UK are taking different approaches to AI, but it would be positive for firms to be able to streamline group activities by using EU approaches in the UK. We note that this will become relevant to data protection 'adequacy' in time, especially if the UK and EU regimes diverge substantially. We strongly support the maintenance of 'mutual adequacy decisions' between the UK and EU. Revocation or invalidation of the European Commission's adequacy decision in favour of the UK would be a significant cost, which would dampen considerably the overall benefits of a more liberal AI regime in the UK.
    b. Given the focus on process and on developers of AI in the EU, as opposed to outcomes and deployers of AI in the UK, direct read-across from one regime to the other may not work. However, there may be opportunities to streamline more tactically. For example, there could be some kind of recognition in the UK that where a firm (or its AI vendor) is applying standards approved under the EU regime, this can be relied on by firms as a strong risk management measure.
    c. International collaboration on the development of global standards and potential governance measures should also form a part of the central function's activities. (See also our comments on the role of standards under Question 21).
    d. International interoperability between varied regimes and approaches should be a topic covered at the forthcoming AI summit being organised by the UK government.
56. Regulatory mapping
    a. The central function may be able to help firms by developing, with regulators, 'maps' of the existing regulatory regime. Such guidance would not necessarily need to be

highly technical but could help firms determine which regimes are likely to apply and direct them to the relevant regulators and regulations.

57. Monitoring public attitudes
    a. Another valuable central function could be research on public attitudes and concerns in relation to AI. Understanding public priorities may help regulators make the difficult decisions about trade-offs between different principles / outcomes when preparing regulation or guidance (eg: between privacy and fairness, or privacy and inclusion). Such research could also inform the central function's educational activities. We do note, however, that research of this kind is already undertaken by the Centre for Data Ethics and Innovation (CDEI), so care not to create duplicative workstreams will be needed. See also our comments under Question 11.
58. Availability of training
    a. Consistent application of regulation can be encouraged by ensuring that appropriate training on technology, regulation and any relevant standards is available. Training is relevant both for regulators and for firms seeking to use AI and apply relevant guidance, including those working in AI procurement within firms. This can help promote a common baseline understanding of the technology, risks and requirements. The central function should monitor the availability of relevant training and could promote its development.
59. As noted under Question 5, government should look at the role of MOUs between regulators to facilitate coordination and the identification of tensions.

Accounting for 'bad actors'

60. The focus of the Whitepaper is understandably on the risks posed by the use of AI to provide goods and services, e.g., the regulation of financial institutions' use of AI. However, there may also be risks associated with the use of AI by bad actors, such as fraudsters or hostile state actors. This risk may not be clearly covered by regulatory authorities, who may be focused on the firms that they regulate rather than the threats posed by outside bad actors against legitimate firms and their customers. Research and intelligence gathering on bad actor AI use can inform guidance and regulation by sectoral authorities, as well as wider policy development.
61. We recommend that the risk and horizon scanning activities within the central function should cover these kinds of cross-cutting and emergent risks in order to help get ahead of them.
62. There should be appropriate connections between the central function and the UK Fraud Strategy and government agencies such as the National Crime Agency and National Cyber Security Centre, as well as security services. Although these entities may not be sectoral regulators per se, they should be included in an appropriate fashion. This will help ensure that they are informed of relevant risks identified by sectoral authorities and will in return help ensure that regulatory authorities are up to date on relevant risks, which may impact the guidance they issue on cyber security and anti-crime measures.
63. The central function – together with NCA, NCSC and other authorities – could also helpfully consider the role AI can play in combating fraud and other crime. This could feed into research notes and regulators' guidance.

Question 11 – Do you know of any existing organisations who should deliver one or more of our proposed central functions?

64. We note at least three organisations that already play an important role in delivering the functions put forward in the Whitepaper:
    a. The CDEI already undertakes some of the tasks identified, such as horizon scanning, gap analysis and risk monitoring. CDEI also possesses – and is developing – expertise in relevant policy and technology areas.
    b. As the ICO has noted in its public response to the Whitepaper, the Digital Regulation Cooperation Forum (DRCF) is also performing risk identification and horizon scanning, and is exploring joint sandbox opportunities among its participants. It also

brings together some of the key regulatory authorities to discuss and work through issues of common interest.

c. Although we are not so close to its work, the AI Standards Hub under the Alan Turing Institute is presumably also engaged already in the design of international standards relevant to AI.

65. Whether or not these bodies are best placed to take on the role (or part of the role) of the central functions in the Whitepaper, government should avoid duplication wherever possible. Some rationalisation may be required, if government intends for one organisation to be expanded to take on the full suite of central functions put forward in the Whitepaper.

Question 12 – Are there additional activities that would help businesses confidently innovate and use AI technologies?

66. As also discussed under Question L1, it would be helpful to develop common reporting, assessment and documentation tools that can be used by AI vendors to show they have taken suitable steps in developing their products. This would be helpful in ensuring firms deploying AI can have confidence that they will be able to use it in a compliant manner. These would not need to be compulsory in the first instance, but could help develop standard practice.
67. These tools would likely be most effective if developed centrally, with sector-specific additions where relevant. We understand the CDEI to be doing relevant work as a part of its AI Assurance workstream.
68. Regular discussions between authorities and industry, as well as consultation papers, would encourage businesses to understand the regulatory limitations, considerations and constraints of innovative AI technologies, along with potential practical applications.
69. Lastly, see also our comments under Question 11 and Questions S1-S4 in relation to sandboxes.

Question 12.1 – If so, should these activities be delivered by government, regulators or a different organisation?

70. As per our previous answer, we see value in central coordination so as to avoid duplication and conflicting measures.
71. On a related point, we query the interaction between the role of the EHRC and other regulators under the Equality Act, noting that all authorities have a duty to prevent discrimination under section 149. This is not an issue unique to AI but given the importance of AI fairness and bias concerns, this overlap in competence should be clarified. We presume that the EHRC leads and that other regulators need to ensure their approaches are consistent with EHRC's guidance, where applicable.

Question 13 – Are there additional activities that would help individuals and consumers confidently use AI technologies?

72. The information provided to consumers when they initially engage with a service or product that involves AI will be important. For example, in some instances it may be appropriate to provide them with information about key potential risks to enable an informed choice (for example, risks to health, safety or fundamental rights). We understand that such information needs will be considered as a part of the ICO's work to produce guidance on the new ADM provisions under the DPDI Bill. Data ethics and 'human design' specialists will likely be able to assist with this work in due course.

Question 13.1 – If so, should these activities be delivered by government, regulators or a different organisation?

73. Please see our response to Question 13.

Question 14 – How can we avoid overlapping, duplicative or contradictory guidance on AI issued by different regulators?

The role of sandboxes

74. Identifying and reconciling potentially conflicting regulatory requirements and expectations could be one of the roles of the multi-regulator sandboxes. We imagine it can be difficult for regulators to always anticipate tensions with other regimes outside of their direct area of competence when this is being attempted in the abstract. Reviewing use cases in the sandbox will bring a practical perspective. (We also discuss sandboxes under Questions 11 and Q12, and Questions S1-S4).

Regulatory coordination and information sharing

75. As noted above under Question 9, we note that regulators must maintain their independence but a convening and coordinating central role can assist in minimising duplication and contradiction. It can also assist regulators by coordinating mapping exercises to identify potential regulatory gaps and areas of overlap, where coordination may be needed (see also comments under Question 10 on regulatory mapping).
76. Increasing understanding between authorities of the core duties of their peers, and the very broad-brush strokes of the rules they enforce, may help each regulator to identify when its work has the potential to intersect with the work of another. It may be valuable to develop a tool such as a checklist to help regulators identify areas of connection, so that they can then consult with the relevant regulator. For example, a checklist could prompt an authority to speak to the ICO where electronic marketing is involved (Privacy and Electronic Communications Regulations), or where there are concerns about the fairness of automating customer-facing decisions (GDPR).
77. Building on this, we note that it may be that tensions occur most often between sectoral regulators on the one hand (e.g. the FCA, Ofcom) and horizontal regulators on the other (e.g. the ICO and the EHRC).
78. Although regulators each need to ultimately make their own decisions, it may be helpful to exercise a 'rule of thumb' whereby a specific regulator takes the lead on certain issues. For example, in the context of the fairness principle it is presumably logical for the EHRC to set the overarching expectations regarding AI, discrimination and human rights, recognising that there will be 'fairness' considerations that are industry-specific beneath this.

Industry consultation

79. Where tensions between regulators are identified, regulators should be encouraged to consult affected industries closely, being the stakeholders most immediately affected. We have noted in the past – not in relation to AI – that where we have called out regulatory tensions, authorities have listened to our concerns but then made rule changes without wanting to first test these with industry. This has at times led to lingering tensions between requirements despite regulators' efforts.

Monitoring and evaluation of the framework

Question 15 – Do you agree with our overall approach to monitoring and evaluation?

80. Overall, we agree, noting that the exact entity or organisation that will take on this function is yet to be determined.

81. In order to be effective and maximise business contribution, the monitoring and evaluation will need to be pragmatic and minimise burdens on firms. In order to be able to provide a helpful response to any data requests, it would be critical for firms to understand how the data will be collected and used.
82. Data collection exercises rarely have good cost/value outcomes if businesses are unclear about how the data will be collected and used. Understanding the specific policy issues the government is trying to address and how the desired data facilitates the monitoring of consumers' interests, would help the industry determine the appropriate granularity, level of data quality assurance and consistency of sources.
83. There will also be data confidentiality issues that will need to be considered with care.

Question 16 – What is the best way to measure the impact of our framework?

84. Consulting / surveying industry will be a necessary part of measuring the impact, for example to identify areas of ongoing uncertainty or tension between regulators' approaches. (See also our comments under Question 15).
85. Similarly, surveys to identify evolving public perceptions of risk, concern and uncertainty in relation to AI would be valuable.
86. It may be appropriate to look in particular at use cases identified as high risk in the UK and also under international regimes, as a point of comparison.
87. We note that one challenge will be in trying to measure outcomes in sectors that do not have a sectoral regulator, as data on regulatory mishaps – particularly lower-level mishaps not triggering public media coverage – are less likely to surface. As such, there is a risk of measurement bias, with a potential to observe more incidents in regulated sectors and to then presume that these sectors are causing more bad AI outcomes, where the numbers could in fact be due to increased transparency about incidents and complaints. International learnings, where AI regulation or sectoral regulation work differently, may be able to provide some insights to fill this gap, as might academic or other research into non-regulated sectors.
88. It may also be valuable to take stock periodically of regulators' fines, sanctions, orders and other enforcement actions involving AI-delivered services, and review any drivers of bad outcomes. Such analysis may help reveal areas where certain risks are falling through the gaps of the framework. (Though noting that regulatory fines can take months or years to be finalised, meaning they will not be an immediate source of intelligence).

Question 17 – Do you agree that our approach strikes the right balance between supporting AI innovation; addressing known, prioritised risks; and future-proofing the AI regulation framework?

89. Yes, we support the approach overall.
90. We note that the success of the framework depends on the effectiveness of the central function. This requires navigating a number of challenges, such as:
    a. Identifying the issues and areas that require effective information sharing and coordination, while avoiding wasting resources or slowing regulatory activities by trying to join up activities that do not require it.
    b. As referred to above in our response, identifying and resolving tensions between the approaches of regulators – or indeed conflicting priorities between them – without encroaching on regulatory independence. These are particularly likely to arise between horizontal regulators such as the ICO and EHRC on the one hand, and sectoral regulators on the other, such as the FCA. There may also be 'territorial overlaps', where two regulators take an interest in substantively the same activity, conduct and risk (e.g., the FCA's Consumer Duty includes guidance on fair use of customer data, despite this matter coming under the ICO's remit as enforcer of the UK GDPR). See also our comments under Question 5 on cross-sectoral guidance and under Question 14.

c. Ensuring regulators are resourced to staff the enhanced coordination and liaison that the Whitepaper framework implies will be necessary. Similarly, regulators will need adequate resources and expertise to understand and manage AI issues. This could require a mix of reprioritisation, new resource from government, shared resource made available via the central function, and secondees from other regulators with more specialised resource.

91. Building on the above: there is a risk under the sector-based approach that we may end up with much more restrictive requirements in some sectors, creating an unlevel playing field that stifles innovation, reduces competition, impacts on the ability to attract AI talent and unduly restricts consumers' access the certain technologies. This is of particular concern for financial services, which is already a highly regulated sector, where an unduly onerous AI framework could largely defeat the 'pro-innovation' sentiment of this proposal. In order to address this risk we recommend:
    a. Ensuring the right level of consistency and harmonisation cross-sector.
    b. Recognising that existing sector regulation that already mitigates AI risks and helps achieve the goals of the AI principles is taken into account.
92. Please also see our comments under Questions 5 and 14 in relation to cross-sectoral guidance and coordination.

Question 18 – Do you agree that regulators are best placed to apply the principles and government is best placed to provide oversight and deliver central functions?

93. Please see comments under Question 17.

Regulator capabilities

Question 19 – As a regulator, what support would you need in order to apply the principles in a proportionate and pro-innovation way?

94. Not applicable.

Question 20 – Do you agree that a pooled team of AI experts would be the most effective way to address capability gaps and help regulators apply the principles?

95. We agree that this would be useful, and should include not only policy and regulatory experts but also technology experts.
96. We also consider that regulators will at times have expertise that could be of value to other regulators, particularly in situations where there is an overlap in competency. For example, ICO has technology expertise, as well as expertise in data processing transparency and AI explanations, which will no doubt be relevant to other authorities' approaches to any transparency expectations they may set. Likewise, the EHRC's expertise and responsibility for discrimination law will be relevant to how others may apply the fairness principle. Secondments between regulators (as well as from the private sector or academia) may be a helpful way to share expertise and achieve alignment of regulatory approaches where this is appropriate.
97. This point clearly overlaps with the point made earlier in our response about managing overlaps in the competencies of different authorities.
98. Separately, we note that the fast pace of development of AI/ML technologies requires constant education, of industry and regulators alike, alongside information-sharing regarding best practices and marketplace developments. We therefore encourage regulators – and the central function – to recognise that engagement and guidance in this space cannot remain static and instead must reflect the dynamism of the technology and opportunities it presents.
99. We also recommend that regulators issue periodic reports identifying capacity gaps that make it difficult for firms in their sectors to comply with regulation, and for regulators to conduct effective oversight.

100. See also our comments on the availability of training under Question 10.

Tools for trustworthy AI

Question 21 – Which non-regulatory tools for trustworthy AI would most help organisations to embed the AI regulation principles into existing business processes?

The role of standards

101. Standards are useful as an optional tool available to firms, as the Whitepaper recognises.
102. Standards are more likely to be relevant to process and governance guidance – such as model risk management – than outcomes guidance, such as fair customer treatment. Overall, we recommend regulation focus on outcomes rather than process but note that there can be a role for both.
103. Where standards are developed, it is important that firms not be required to use specific standards. This is not consistent with an outcomes-based approach to regulation; there can be multiple legitimate means by which key outcomes are achieved. In sectors such as financial services, where there is already an extensive set of rules and significant corresponding compliance burden, forcing the adoption of specific standards would add to firms' compliance burdens unnecessarily.
104. Furthermore, standards can take a long time to update and may not be fully applicable to emerging use cases or techniques.
105. There is nonetheless a place for recognising standards and specific governance controls (e.g., incremental model rollout or 'circuit breakers') as being potential steps supporting compliance with regulatory requirements. Adherence to certain standards could be recognised as amounting to a risk mitigation measure or as showing compliance with relevant regulatory requirements. This would encourage uptake and participation in the development of standards. (As noted above, this is likely more relevant to process and governance than to outcomes-focused regulation).
106. We note, however, that this would ultimately be a decision to be made by each regulator.

Further considerations relating to standards

107. It could be beneficial from the perspective of international interoperability to recognise that conformity assessments and adherence to standards under the EU's AI Act can provide assurance (or a partial assurance) to firms that an AI product being procured is compliant with relevant requirements.
108. See also our comments on AI assurance tools in the AI supply chain under Question 12 and Question L1, and our comments on the central function's role under Question 10.

Final thoughts

Question 22 – Do you have any other thoughts on our overall approach? Please include any missed opportunities, flaws, and gaps in our framework.

109. Government and regulators share an important role in helping the economy to maximise AI's benefits while managing the risks of the technology. Not all new AI use cases will be high risk and it is important for all stakeholders to acknowledge this. The explainability of the model or system combined with nature of the product or service incorporating the AI will likely be key factors.
110. That said, we do note that there may emerge use cases over time which lack a sectoral regulator, but which do give rise to high risks. In this case, it may be that cross-cutting regulators such as the ICO, EHRC and CMA are able to cover the risks, but this will require ongoing monitoring. There are a number of potential gaps that we note so far, as set out below.

HR and recruitment use cases

111. These are high impact but lack a clear sectoral regulator. As above, it may be that the EHRC and ICO are well placed to cover the key risks, but this will require consideration.

Digital platforms

112. There may be a gap around digital platforms or other tech firms that are not subject to the CMA's Digital Markets Unit due to falling below the 'strategic market status' threshold, and similarly fall outside of the 'online harms' regime's scope.

Generative AI, openly available to the public

113. Over recent months these kinds of applications have attracted considerable media attention and clearly offer great potential benefits to consumers and businesses.
114. A risk-based approach ensures that AI regulations remain relevant over time and remain proportionate to the risks posed by the application. This assessment should largely focus on the outputs and potential impacts of the application – for example lending, fraud detection, ranking social media posts or online search – irrespective of the exact technology powering them.
115. Generative AI systems, like many AI tools, can be used in a wide variety of downstream applications, with very different types and levels of risk. In many cases, it may be difficult to implement effective mitigations against risks at the model level. Effective risk management will require evaluating the risk of deploying generative AI in specific applications for specific use cases and user bases, and tailoring appropriate risk mitigations to those circumstances.
116. Insofar as generative AI is deployed in regulated sectors, the existing regulatory regimes – such as the model risk management framework apply to financial services firms – should largely address key risks already. The deploying firm is responsible for any unfair bias or other harms, and for ensuring there is adequate transparency to answer regulators' or consumers' questions. As a starting point, firms will need to assess whether a LLM (or other form of generative AI) is the right model for their use case.
117. For their part, providers of general-purpose generative AI should ensure that their technology meets a high bar for performance and general safety. If the technology is intended to be used in high-risk applications, the developer should provide appropriate transparency and documentation to those firms deploying the technology to enable them to effectively manage risk. (See also our comments on AI supply chain issues under Questions L1 to L3).
118. Notwithstanding the above, the key novel challenge with these prominent new products may be their freely available nature. The availability of these tools to the broader public raises at least four challenges:
    a. In some cases the AI tool developer provides services to the public. As such, the developer would be responsible for consumer outcomes, but it may not in fact be aware of exactly how the tool is being used by each consumer and what requirements therefore apply. Furthermore, the consumer may not be aware of the tool's limitations, such as inaccuracy, and therefore utilise it in inappropriate ways.
    b. The AI tool may be used by companies for purposes not anticipated by the developer and without a normal procurement and due diligence process. The user company might also lack the expertise to identify and address regulatory or ethical requirements.
    c. The AI tool could be used by bad actors, eg to create misinformation, to generate materials for defrauding consumers, or to defeat legitimate firms' customer authentication tools (for example by generating false documents or photos, or credible imitations of customer voice or other markers).
    d. We further note that – where a tool is publicly available on the internet – international cooperation will be crucial to addressing the associated risks effectively.
119. Such services warrant accelerated and focused attention from policy makers and regulators in order to review these – and other – risks and challenges.
120. In relation to LLMs more generally, whether freely available online or not:

e. The use of LLMs trained on public data sources has also been acknowledged to pose data protection and privacy challenges. It may be too late to 'put the genie back in the bottle' on this issue, however.
f. Foundation models can be deployed as part of an AI supply chain, like other AI tools, which add to the difficulty in identifying, analysing and addressing risks. Human oversight and review of generated outputs will therefore likely play a critical role in the responsible deployment of LLMs.

Keeping pace – regulatory perimeters and existing statutes

121. We also note that certain products or services may emerge which resemble in substance regulated activities but fall outside of the technical scope of regulation. A recent example in financial services is 'by now pay later', for which the FCA's scope has needed to be extended. The central function will need to monitor for such developments.
122. Finally, we note that existing statutes may turn out to be ill-suited to certain use cases or models that emerge over time, in the sense that they might unduly inhibit innovation. This risk should be monitored for by government and the central function. We note that the post-DPDI Bill provisions on ADM allow for statutory instruments to grant exemptions and refine the safeguards, when required. These are helpful future-proofing measures, though they should be used only following full public consultation and carefully cost-benefit analysis.

Legal responsibility for AI

L1. What challenges might arise when regulators apply the principles across different AI applications and systems? How could we address these challenges through our proposed AI regulatory framework?

123. In the context of a firm using an AI tool provided by a vendor, we note that the entity legally responsible in the event of an incident will vary depending on the relevant legislation and regulatory regime. For example, for a data protection law breach, the liable party will generally be the data controller. This would often be the firm deploying the AI tool but could simultaneously include the vendor as well, if the vendor has a high level of control over the nature of the data processing. However, which firm or firms is/are a data controller is a question of fact, not contract, and can at times be uncertain.
124. In contrast, in terms of financial services requirements, it will invariably be the financial service provider that is responsible. Firm responsibility for customer outcomes is a key tenet of financial sector regulation.
125. Furthermore, and as discussed above in relation to LLM challenges, where a model is used to do analysis at a macro level, with the outputs then utilised by other firms to provide customer-facing services or products, determining the liable party may be a challenge. It may not be clear to what extent, or in what contexts, the firm using the outputs of the initial macro model should be able to rely on those outputs.
126. The difference in responsible parties under different regulatory regimes, combined with the challenges stemming from AI supply chains, will bring complexities, though contractual indemnities can play a role in resolving these. Exploration of best practice or optional 'standard contract clauses' might be helpful here, as would a mapping exercise by the central function to develop a common understanding of the placement of liability in different contexts.

Vendor assurance

127. We also note that the firm deploying a bought-in AI system will need adequate information to satisfy itself that its use of the system will be compliant with the relevant laws and regulations. However, there are legitimate reasons for which AI vendors may be reluctant to share all information with clients that might be asked for. This can raise tensions between the two parties, which need to be balanced:
    a. On the one hand, the client firm will need sufficient visibility of the workings of the tool (transparency, explainability) to stay confident about ongoing compliance. (In addition

to the need for effective internal processes such as staff training, monitoring, issue escalation, reporting and remediation). Where a model is developed by a developer, it may be fine-tuned by another, and then deployed by yet another actor down the chain. If a problem such as bias exists in the upstream model, then the downstream model will be affected also. The impact will likely manifest or become apparent in the downstream model, but it may be difficult to determine the root cause. Firms need an appropriate level of visibility of the design / theory / logic of the model, and of control and governance processes.

b. On the other hand, demands for detailed access to vendor model development features and internal data can raise vendor concerns about intellectual property, security and model integrity (especially with adversarial models). There is a risk of disincentivising model innovation if too much information sharing is expected. In some instances, deploying firms may seek access to information or data that would bring risks for the vendor, without materially benefiting compliance processes. Information needed to demonstrate compliance in relation to simple linear regression models will not be identical to what is needed in relation to a complex machine learning model, for example.

128. A number of measures may help bridge this gap, for example:
    a. Development of 'targeted use' classifications for models, clarifying the intended use cases for a given AI product.
    b. Use of 'model cards', setting out information needed for the deployer to responsibly make use of the model. (The model card would need updating if the deployer were to retrain the model or deploy it for a new use case).
    c. Regulatory guidance on the risk levels of different categories of use cases – potentially with indicative chief risks – could help developers and deployers of AI systems to have a shared understanding of regulatory expectations regarding risk and risk mitigation, and the corresponding information needs of the deploying firm.
129. These kinds of information challenges between technology vendors and firms deploying technology are not unique to AI, being a factor in wider procurement and contract negotiation. However, there is **merit in promoting the development of optional AI assurance techniques**, as the CDEI is already doing, and in monitoring for any potential market failures that might impede AI developers and AI deployers from finding common ground on information requirements in AI procurement.
130. There may be a particular challenge in the context of opaque of complex tools that are widely available to relatively small or unsophisticated deployers. The regulatory risk may sit with the small deployer, which may lack the technical expertise to manage it, and might conceivably lack negotiating power to obtain the necessary documentation needed to do full due diligence, if indeed such a market failure develops. This may be particularly relevant to generative AI tools, especially when these are made readily available online.
131. See also our responses to Question 12 and Question 22.

<u>Managing sectoral differences at vendor level</u>

132. As we have stated, we support the sectoral approach. However, we note that at the vendor level this could lead to complexities where substantively the same product is provided into multiple sectors. Where the regulatory expectations of each diverge, it is possible that vendors will need to produce multiple versions of the same product, tweaked to accommodate the varying sectoral regulatory approaches. There is a risk of unnecessary duplication, though we do note that divergence between sectors will sometimes be appropriate.
133. See also our comments under Question 5 about collaboration on cross-sector guidance.

<u>International considerations</u>

134. Lastly, we note that the EU is working on an AI liability directive. It may be that alignment with the EU approach will be necessary over time. The approach to liability taken by the US and Asia-Pacific countries will also need monitoring.

L2.1. Do you agree that the implementation of our principles through existing legal frameworks will fairly and effectively allocate legal responsibility for AI across the life cycle?

135. Please see previous answer and our comments under Question 4.

L.2.2. How could it be improved, if at all?

136. Please see comments above under Question L1 in relation to vendor assurance.

L3. If you work for a business that develops, uses, or sells AI, how do you currently manage AI risk including through the wider supply chain? How could government support effective AI-related risk management?

137. Some of our members do operate as AI developers. Please see our comments under Question L1.

Foundation models and the regulatory framework

F1. What specific challenges will foundation models such as large language models (LLMs) or open-source models pose for regulators trying to determine legal responsibility for AI outcomes?

138. We presume that this question is largely intended to cover publicly available generative AI. Please see our response to Question 22.

F2. Do you agree that measuring compute provides a potential tool that could be considered as part of the governance of foundation models?

139. Compute capacity would be one relevant benchmark. This work by the OECD should prove a useful reference: https://oecd.ai/en/wonk/ai-compute-capacity.

F3. Are there other approaches to governing foundation models that would be more effective?

140. As covered in our response to Question 22 (and Question L1 in relation to AI supply chains), we note that several challenges exist in relation to LLMs and generative AI. Although existing regulation should be effective in regulated sectors, and controls need to be proportionate to the risk of the each *application* of the model, the openly available some relevant AI tools, and their international nature, could prove to be a gap in the proposed approach, warranting priority attention from policy makers.

AI sandboxes and testbeds

S1. To what extent would the sandbox models described in section 3.3.4 support innovation?

141. We particularly support multi-regulator sandboxes that prioritise use cases with the potential to create tensions between different regulators' priorities. Putting such use cases into a multi-regulator sandbox will help to expose – and ultimately resolve – areas of tension.
142. See also out comments under Questions 11, 12 and 14.

S2. What could government do to maximise the benefit of sandboxes to AI innovators?

143. Mechanisms will be needed to ensure that the right regulators are identified for participation in relation to each use case taken into the sandbox. It would be a missed opportunity – and give misleading outputs – if a relevant regulator were not included in a case that touches on their domain.
144. Sandbox reports will need to be very clear about which laws and regulations were considered, and which were not.
145. Though not necessarily forming part of a sandbox per se, it would be helpful for government and / or regulators to support the development of technical metrics and shared test sets in order to create common benchmarks for testing. Such technical metrics and testing benchmarks could be used to assess whether there is proper alignment between model output and business goals and also act as a control on whether data quality issues exist.5 We note that the FCA is already investigating the potential role of synthetic datasets.

S3. What could government do to facilitate participation in an AI regulatory sandbox?

146. Government can actively promote participation and sponsoring of the sandbox to enable regulators better to add value for industry.

S4. Which industry sectors or classes of product would most benefit from an AI sandbox?

147. As touched on above, industries and use cases that touch multiple regulatory / legal domains would be the most helpful to cover.
148. High risk or complex use cases would also benefit most from access to sandboxes. This could include use cases with complex considerations relating to bias and discrimination, such as the boundaries of 'positive action' version 'positive discrimination' in efforts to improve model fairness.

ENDS

[5] An example of such technical metric standards development that can provide helpful clarity to stakeholders is NIST's Face Recognition Vendor Testing program, available at https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-ongoing.