AI Foundation Models: initial review
Consultation by Competition and Markets Authority

Written evidence submitted by the Trustworthy Autonomous Systems Hub.

Response Authors:

Dr. Joshua Krook: Research Fellow in Responsible AI at the University of Southampton. He has a PhD in Law from the University of Adelaide and previously worked in technology policy for the Australian government.

Professor Stuart Anderson: Professor of Dependable Systems and Co-I TAS Governance and Regulation Node, The University of Edinburgh.

Dr. John Downer: Co-I Functionality Node and Senior Lecturer in Risk and Resilience, School of Sociology, Politics and International Studies (SPAIS), University of Bristol.

Dr. Peter Winter: Research Associate in Regulation of Autonomous Systems, Functionality Node, School of Sociology, Politics and International Studies (SPAIS), University of Bristol.

Professor Derek McAuley: Professor of Digital Economy in the School of Computer Science at the University of Nottingham, Director of Horizon, fellow of the Royal Academy of Engineering.

Professor Sarvapali D. Ramchurn: Professor of AI and Director of the UKRI Trustworthy Autonomous Systems (TAS) Hub. Sarvapali has over 19 years of experience in developing AI solutions.

Citation
Trustworthy Autonomous Systems Hub. (2023) Response to: AI Foundation Models: Initial Review, Competition and Markets Authority. https://doi.org/10.5258/SOTON/P1110

Contact details:
J.A.Krook@soton.ac.uk

About the TAS Hub:

The UKRI TAS Hub assembles a team from the Universities of Southampton, Nottingham and King's College London. The Hub sits at the centre of the £33M Trustworthy Autonomous Systems Programme, funded by the UKRI Strategic Priorities Fund. The role of the TAS Hub is to coordinate and work with six research nodes to establish a collaborative platform for the UK to enable the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice by individuals, society and government. Read more about the TAS Hub here.

Question 1: How the competitive markets for foundation models and their use could evolve

Given the current market conditions, it is unlikely that a flourishing “market” of foundation models will develop with robust competition.

Foundation models are currently concentrated and controlled by a small number of international companies with the capacity and resources to afford the computing power necessary to create them. The costs associated with software development, training and maintaining foundation models is currently extensive, meaning that it is currently difficult for smaller companies to compete or enter the market. A truly competitive market for foundation models will likely require some form of intervention regarding the cost of training and maintaining these systems. However, a growing community around open source software, models and data assets may be a route via which the technology can be democratised and enable smaller organisations to collaborate to enter the market with less individual investment.

It is important to note that many of the large companies currently either engaged in the foundation models work, or funding such work, are those who would be deemed already to have Strategic Market Status in various markets, and can be seen to be already leveraging their domination and monopoly revenues in other areas, to skew the market. Indeed, there is substantial lobbying effort being deployed via the rhetoric of existential threat to secure a regulatory position where only companies that can be fully vertically integrated can enter the market.

Such companies already offer API services to other businesses, which might be seen as a route to a competitive market in applications of such technology. However, as we have seen in many other digital platforms, the platform provider will also enter the market for applications, but at a marked advantage as they have visibility of training data and model, as well as competitors use of the API on which to build, and can cherry-pick the most lucrative markets. Rapid action at this point to establish an appropriate Code of Conduct is preferable to post facto remedies.

Question 2: What opportunities and risks these scenarios could bring for competition and consumer protection

There may be a “secondary” market that develops in relation to tools to adapt foundation models for particular contexts. Even in this secondary market however, it is likely that the major tech companies will exercise a significant degree of control and impinge on market competition.

Foundation models are likely to be incorporated into many products so there is also a risk of some level of market power for the creators of foundation models across a wide range of digital products.

There is also a significant risk of “common mode failure" if a foundation model has some significant flaw. Common mode failure refers to a situation where two or more components of a system fail in the same way at the same time, leading to potentially wide scale disruption of the system. If a foundation model is incorporated into a variety of products and then experiences common mode failure, this could incur significant financial cost and other foreseeable harms on the market.

It is clear from the responses of many of the existing Large Language Models (LLMs) that they have ingested significant quantities of personal information and copyright material. As such, consumer and arts rights, including copyright, may have already been breached. There is a need for transparency around the data and code used for training to ensure such rights are protected; while for confidentiality reasons this may not result in full public transparency, it must include the ability of regulators to investigate the system’s workings akin to the powers within the proposed EU Artificial Intelligence Act (European Commission, 2021).

However, some transparency is also important to those building applications via APIs to ensure that they can meet their obligations. In some markets and sectors this can be considerably more complex (e.g. Financial Services) than will have been considered in training of the underlying model. In other words, further obligations may be relevant for secondary markets that extend the API functionality or application.

Question 3: Which principles can best guide the ongoing development of these markets so that the vibrant innovation that has characterised the current emerging phase is sustained, and the resulting benefits continue to flow for people, businesses and the economy

Control over any primary market in foundation models by a national body is unlikely to succeed. There should be a focus on providing good regulation for the development of tools for the specialisation of foundation models to specific contexts and on ensuring good control over the data used in specialising the models to specific contexts.

The unique properties of foundation models that make them attractive as general-purpose AI may also present unknown and unpredictable risks. The issue of 'stepping in to address risks when necessary' may actually come too late.

A first step to the governance of foundation models would be the support and reinforcement of the Digital Markets Unit (DMU) to effectively govern and regulate foundational AI models, including predatory practices by major firms. The DMU should be encouraged to, in effect, create and implement rules for different risks (e.g., disclosure around data being used, performance, compute) and require companies to show their work. While existing regulatory practices are already in place and span multiple domains, these bodies are under-resourced and have failed on many occasions, especially around matters of data privacy in the digital sector and Big Tech (Edwards, 2022; Garrod et al., 2023). The DMU should be supported with greater resources and research. For example, this may take the form of pre-deployment and post deployment testing, as well as identifying/making sense of bad actors and all sorts of risky behaviour.

A second step highlights a need for the DMU to demand transparent development of foundation models. While this isn't new, the promotion of transparency in the development process is a good first step in the governance of foundation models. This would be, for example, ensuring foundational AI companies have mechanisms in place to openly share information about the model's training data, architecture, and potential biases, which opens them up to scrutiny through other actors (such as external auditors or external researchers). Doing so will help facilitate better understanding of the model.

A third step would be to establish ethical guidelines for the development and use of foundation models. Such guidelines should be specific to foundational models and address issues such as fairness, privacy, security, and the avoidance of harm.

A fourth approach would be on interdisciplinary collaboration and multistakeholder involvement in the co-production of rules and ethical principles. This could involve a diverse range of stakeholders, including researchers, policymakers, industry experts, and the general public all being involved in shaping these rules (For instance – through advisory boards, expert panels and/or public consultations).

A fifth approach would focus on robust evaluation and testing. Bommasani et al. (2022: 17), in particular, points out how foundation models challenge the existing standards of contemporary evaluation paradigms in machine learning since they are "one step removed from specific tasks". For this reason, Bommasani et al. (2022: 17) endorse the creation of three new rigorous

evaluation processes to assess the performance and potential biases of foundation models. Through three central nodes of analysis, Bommasani et al., (2022: 17) emphasises: (1) a process which evaluates foundation models *directly* to measure their *inherent capabilities* as a means to inform how foundation models are trained ("intrinsic evaluation"); (2) a process which evaluates task-specific models by *controlling for adaptation resources and access* ("extrinsic evaluation and adaptation"), and (3) a process which supports a broader *evaluation design* to provide richer context beyond measures of accuracy (e.g., robustness), fairness, efficiency, environmental impact ("evaluation design"). The creation of these three evaluation processes and testing infrastructures give hope to AI companies as they identify and mitigate biases in foundation models, especially as they look to address questions of fairness across different demographic groups. For example, adopting a process of 'intrinsic evaluation' could lead to the development and use of debiasing techniques which actively diversify the training data which, in turn, can be used to evaluate disparities in performance. In connection with that promise and against the exacerbation of unfair outcomes that arise from foundation models, Snorkel AI's data-centric platform 'Snorkel Flow' is intended as an important contribution to the identification and management of biases in inherited foundation models, with the aim of "correcting biases in AI systematically".

A sixth approach to the governing of foundation models could be the creation of accountability mechanisms. Defining clear lines of responsibility and accountability for the development and deployment of foundation models is especially important given that foundation models (by definition) are incomplete, but can be adapted for use by an AI user (like an insurance company or telecommunications company) across different domains like industry, science, government, and academia. This could involve mechanisms for reporting and addressing any concerns or complaints raised by AI users, third parties or affected communities.

A seventh approach to governance should focus on applying updates (Dai et al., 2021) or learning such update rules (Mitchell et al., 2021). This updating and improvement of foundation model should incorporate feedback from users and stakeholders; an iterative design process which should help to ensure that the model evolves with societal needs and values.

References:

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E. and Brynjolfsson, E. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

Dai, D., Dong, L., Hao, Y., Sui, Z., Chang, B., and Wei, F. (2021). Knowledge Neurons in Pretrained Transformers. Available online: https://aclanthology.org/2022.acl-long.581.pdf [Accessed: 05/05/2023].

Edwards, L. (2022) 'Regulating AI in Europe: Four problems and four solutions'. Available online: https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf [Accessed: 18/05/2023].

European Commission. (2021) Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 [Accessed 02/06/2023].

Garrod, D., Pettifor, S., Meriani, M., Chinoy, K., Prota, M., Lahtinen, T (2023) 'Big tech's Growing Regulatory Burden in Europe – Failing to Prepare is Preparing to Fail'. Available online: https://www.akingump.com/en/insights/alerts/big-techs-growing-regulatory-burden-in-europefailing-to-prepare-is-preparing-to-fail#authors [Accessed: 18/05/2023].

Mitchell, E., Lin, C., Bosselut, A., Chelsea Finn, and Manning, C.D. (2021). Fast Model Editing at Scale. In International Conference on Learning Representations. Available online: https://openreview.net/pdf?id=0DcZxeWfOPt [Accessed: 06/05/2023].

Snorkel AI (2023) Weak Supervision. Available online: https://snorkel.ai/weak-supervision/ [Accessed: 02/05/2023].