POLICY on ARTIFICIAL INTELLIGENCE (referred to as AI throughout for ease of reading)

BACKGROUND

AI (also known as Machine Learning, Natural Language Processing, Large Language Models) means any device or linked devices that use digital processes on any inputs to produce outputs similar to or better than those that a creative, literate, numerate and well-informed human might give, for similar purposes e.g to persuade, solve, synthesize, investigate, correlate, make decisions, optimise processes etc. Current examples include but are not limited to: recommender/influencer software, algorithmic trading or commercial tools, security and defence systems, data processing or harvesting, biometric and healthcare measurement, scientific research tools, generative AI software such as chatbots, neural networks, customer interfaces. The scope to include new uses for digital technology as it emerges into this definition must remain in perpetuity, in order that these may be subject to necessary controls.

The development and widespread use of AI within this meaning has clearly already begun, but there is no common agreement about the exact point we are at, the rate of direction of change, nor whether progress will be smooth or marked by sudden explosive periods. It is definite though that there is no practical way of turning back the clock and "un-inventing" AI. In a similar way to the emergence of nuclear technology in the 20th century, there are possible ways AI may develop to be extremely good for humanity and the environment and others in which it could do massive harm. Consequently, we need a system that steers development along benevolent, sustainable pathways though a combination of regulation, guidelines and checks for compliance which would embody the Precautionary Principle to which we should adhere "when scientific measures are uncertain and the stakes are high" (EU Thinktank).

PRINCIPLES

We echo the existential threat statement, signed by many experts worldwide[1], issued 29/5/23:
"Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."
To address this we propose below
- policy statements on fostering international co-operation towards the establishment of a world-wide body similar to the International Atomic Energy Agency that could oversee and police AI development globally.
-policy statements on the creation of an independent body of experts to monitor compliance with proposed regulation of AI, and the power to enforce the banning or temporary shutdown of AI systems if dangerous practices arise from their use.

AI development should be based on ethical principles for development and deployment so that systems using AI are safe, robust, transparent, accountable, and human-centric. AI developers should aim to produce systems that are socially and environmentally responsible, that include safeguards against inappropriate or under-age use, measures to address and rectify existing bias, discrimination and equality issues, measures that protect individual privacy and liberty, ways to rectify mistakes made and ways to allow use to be switched off or paused. Despite the best intentions of regulation, the multi-faceted nature and complex abilities of AI will create scenarios which prior testing cannot foresee, so the use of AI should as a matter of routine include regular monitoring and reportage of unanticipated outcomes for independent expert review.

New AI developments should be shared as widely as possible and for the benefit of all. Any research conducted by public institutions (universities etc) should become open for access by all.

The burgeoning uses of AI within the commercial sphere should be required to put social and environmental priorities ahead of financial returns to shareholders.

The use of AI in law enforcement and judicial matters should be facilitated but carefully monitored, including assessing the impact on fundamental rights prior to and during the deployment of AI-enhanced systems, and specialised training for police officers and judicial personnel concerning the ethical and safe use of such systems.

The use of AI across the healthcare sector and into wider society should also be facilitated but carefully monitored, including clear guidance on the levels of human oversight in decision-making and training for healthcare workers concerning the ethical and safe use of such systems.

In the UK AI development should aim to cater for all basic needs, freedoms, personal safety and maintenance of physical health for every individual to a standard similar to that enjoyed in G7 countries today. AI development and usage should then encourage and facilitate for all individuals the freedom to live a worthwhile life, devoting time in whatever proportions they choose to the following "seven C's": - Curiosity, Conservation, Challenge, Creativity, Community, Charity, Care.

AI will create a new pattern of distribution of wealth and earning power with the potential for a few to accrue a great deal while many lose out, so we should ensure our tax policies are able to address this, with the aim that taxation could offset the costs of a Universal Basic Income scheme in the long term, in the shorter term could be used to enable furlough-like payments to the many whose livelihoods may be destroyed by AI.

There are several technologies based on AI whose evolution is fluid and fast which pose significant challenges to effective regulation while being open to pernicious use. These include social media influencer programs, news media, advertising targeting, blockchain information or transaction processing, "virtual friendship" apps, visual enhancement/ deepfaking apps, and others. So the first task of a regulatory body we shall create will be an immediate review of the current usage of these with the power to restrict or prohibit their further use.

We assert that every individual has a right to the ownership of their own identity in terms of appearance, sound and creative output. So the use of AI to mimic, impersonate or create likenesses of any individual for commercial, persuasive or other reasons will be prohibited (with the possibility of exceptions being granted in certain cases). This right cannot be over-ridden by an individual signing a waiver, for example an actress signing such an agreement that her image/voice/technique could be artificially manipulated in future to create new material seeming to come from her.

The use of AI in advertising represents a direct challenge to the need for the management of consumption demand in the face of biodiversity and climate crises, with the potential for health and wellbeing issues to be created or exacerbated by advertising. So in addition to regulation the creation of an independent body of experts to monitor compliance with proposed regulation is needed in this area.

OBJECTIVES

In the short term we seek to direct the pathway of development for AI and its uses through a combination of regulation, risk categorization, guidelines and compliance checking, while enabling and encouraging growth, innovation and research in the many possible beneficial uses of AI.

In the long term we seek ever-improving goal-alignment of AI systems to those outcomes which are beneficial to protect humanity and its freedom, the environment and its sustainability.

SHORT TO MID-TERM POLICIES

(Policy statement on banned AI uses or developments)

DCMS 101 To protect current freedoms and democratic processes, to protect personal privacy and data, also to prevent discriminatory or biased practices from propagating, new laws will be created prohibiting certain uses of AI. These include, but are not limited to:- continual biometric identification in public spaces; social or healthcare scoring systems; autonomous weapons systems; "deepfaking" or other image generation to manipulate or control opinions; data-mining, analysis or storage contravening current UK GDPR regulations; advertising and other uses that seek to target children or vulnerable people inappropriately.

(Policy statement on risk categorization of AI technologies in line with EU AI Act)

DCMS 102 A new regulatory framework backed by legal enforcement will be created in which certain uses of AI or AI packages will be categorized as "High-Risk". This will be decided in terms of both their design and their intended use, with the High-Risk category being applicable to any AI use that may significantly affect the lives of individuals in ways that might not be in their best interests. These will be subject to stringent controls in advance of and during their release. This categorization will be applied in line with the coming EU AI Act[1]. Other AI uses or systems will not be subject to such stringent regulatory requirement but there will be provision to move any existing or new AI technology into the High-Risk category if the need arises and suspend or prevent its use until it conforms.

(Policy statement on encouraging innovation subject to compliance checks)

DCMS 103 Compliance checking for the developers of AI needs to be streamlined to encourage innovation. A common procedure such as the CapAI system[2] will be used to effect this.

(Policy statement on monitoring AI nationally and internationally)

DCMS 104 Working with international partners we shall push for the creation of a single independent body monitoring AI development with the aim of preventing existential threat or other inappropriate use, development or deployment of AI, in a similar way to how the International Atomic Energy Agency currently monitors nuclear technologies world-wide.

DCMS 105 In the UK we shall create a new independent body to monitor and categorize AI development, deployment and use and to ensure that regulatory framework and other guidelines on use are followed. This will replace the current patchwork of responsibilities shared by Ofcom, ICO, etc which allows too many loopholes or areas where it is not clear to whom responsibility falls when a threatening use arises.

 (Policy statement on right to redress)

DCMS 106 Within regulation there will be a means for individuals to seek redress or explanation of AI-based decisions, necessitating that AI systems have the ability to trace back through a decision-making process in a way that is clear to humans.

(Policy statements on need for AI transparency, identification to end users, traceability and elimination of bias.)

DCMS 107 Within AI regulation will be the requirement for AI to identify itself to users. If a person is in a conversation with AI it will identify itself in a way that is clearly understood by the person. If a person receives diagnosis, sentence, result etc that has been generated by AI it will be made clear to the person how and to what extent AI has been involved in the information given and how to challenge this.

(Policy statement on equality, disability, discrimination and regional disparity)

DCMS 108 AI technology represents a significant opportunity to address areas of inequality of access, discriminatory practice or regional inequality of opportunity both in existing and future society, with its ability to compare large amounts of data, identify trends and suggest solutions. We shall encourage and enable this through sponsoring research.

(Policy statements on AI in Security, Peace, Defence and Law Enforcement to clearly delineate acceptable uses for AI and possible exception scenarios.)

DCMS 109 In the interests of security, national defence and the maintenance of peace there are possible AI-use scenarios that should be planned away in advance. In addition to the banning of "deepfaking" or other image generation, there is the possibility that in other ways social media using AI might become a source of misinformation or subversive, inflammatory material. These will be prevented both by the monitoring body we shall create but also it will become a requirement of the owners of social media platforms that they have effective measures against such use.

DCMS 110 Also in the interests of security and national defence the monitoring of private communications is often desirable, but this must be balanced against both the right to privacy and the need to protect effective encryption so that secure communications and transactions might take place. So within regulation there will be a mechanism by which security agencies can apply for judicial permission to monitor communications in a selective way.

DCMS 111 In a similar way to the previous statement, there may also be times when real-time biometric identification of individuals in public becomes desirable, for example to counter imminent terrorist threat, and again for this purpose security agencies may be able to seek judicial permission to do this.

DCMS 112 We will not allow our security or defence forces to use AI-driven autonomous weapons and targeting systems, ensuring that "human-in-the-loop" control is always required, and work internationally with other agencies to ensure this is the case world-wide and thus prevent an AI arms race from developing.

DCMS 113 In the long term one of the goals that AI will be aligned to is the reduction of tension between nations or factions within them, progressively reducing our dangerous build up of lethal technology and the anguish, inhumanity and waste of resource it represents, until world peace is established and maintained.

(Policy statements on taxation and the economy).

DCMS 114 Our existing fiscal policies allow scope for suitable taxation of those who will accumulate wealth either directly or indirectly as a result of AI, and these will be reviewed and revised if the distribution of wealth and earning power shifts.

DCMS 115 Although we recognize the many concerns over jobs and livelihoods owing to AI development, we also see the growth of AI as making more immediate the need for Universal Basic Income, which we will introduce (current policy EC402). Far from being just a safety net, the overall aim of this would be to enable people to break free from the current model where so many labour essentially to benefit already wealthy shareholders and propagate our destructive levels of

consumption. Universal Basic Income we see as a starting place for allowing us all to develop harmonious lives in which we feel enriched and valued beyond our current market-orientated measures; this truly is the potential of AI to transform our society and individual experience for the better, which we strive to bring into being.

(Policy statements on AI in media and culture specifically citing areas to regulate, need for authentification of news sources, need for intellectual property protection and areas to encourage.)

DCMS 116 In the context of media particularly news/current affairs it should be clear that the purpose of using AI should not be to further sensationalise or even invent content, nor to demonise groups or exacerbate divisions between groups in society. As well as regulatory control since these uses will be categorized as High Risk, the controllers of news media will be required to develop a new code of practice and police this between themselves.

(Policy statements on AI usage for health and wellbeing, specifically citing areas to regulate and areas to encourage)

DCMS 117 The uses of AI in all aspects of health and wellbeing is one of the most established current areas of its development. By classification as High Risk we will regulate the use of healthcare data in a manner that is safe and transparent to all users and regulate to ensure that end users are aware how and when AI is being used in making healthcare decisions

DCMS 118 We recognise that the use of AI in most aspects of healthcare is one of the most immediate areas in which the positive aspects of AI will be seen, also it is a way in which the health equality divide referred to in the Marmot Review report may be addressed. So we will encourage and facilitate:- Making routine the use of AI to augment (**not** replace) healthcare professional judgement, for example by concurring with diagnoses and prescription, checking decisions against the most up-to-date research (but still subjugate to Doctor's final oversight and decision, using a "trust, but verify" model); making routine the use of AI to save paperwork time e.g transcribing consultation notes, producing written communications between healthcare providers, discharge notes, notes to employers etc, with the aim of allowing more 1:1 time with patients and reducing overall workload and stress; Creating individualised patient treatment plans and carrying out many of the routine management tasks of these.

DCMS 119 We will work with NICE, the NHS and other stakeholders to ensure we maintain and grow the current framework by which the many individual research projects on using AI in particular health areas are currently funded[4] .

DCMS 120 In addition to the above statements which are beneficial from the standpoint of established healthcare providers, the benefits for all members of society will be maximised by creating freely available training on using AI to coach us all in healthy lifestyle plans thus obviating some of the need for later health care.

(Policy statements on AI in educational context, specifically citing areas to regulate and areas to encourage)

DCMS 121 Within educational contexts we shall regulate the use of AI data processing for assessing individual students or educational institutions of all sizes in a manner that is safe, ethical and transparent and also regulate to ensure that end users are aware how and when AI is being used in making educational decisions.

DCMS 122 We shall work with Ofsted and other stakeholders to offer guidelines for declaring how and when AI usage needs to be acknowledged by students in the production of work, to protect against plagiarism and allow true and reliable assessment of abilities.

DCMS 123 In educational contexts the ways in which AI could be beneficial will also be guided in collaboration with Ofsted and others, including making routine the use of AI to augment (**not** replace) pedagogical professionals at all levels, including training them in how to teach students the skills required to make positive use of AI in their subject area or context. We also aim to make routine the use of AI by pedagogical professionals to save paperwork time e.g transcribing meeting notes, producing written communications between institutions, reports, references etc, with the aim of allowing more time developing learning, additionally reducing overall workload and stress. Additionally, we will facilitate the use of AI in creating individualised learning plans and carrying out many of the routine management tasks of these, acting as friendly guide and coach to enable students to stick to them.

DCMS 124 In addition to the above statements which are beneficial from the standpoint of established educational institutions and students, the benefits for all members of society with the aim of equalising access to the highest quality educational content and guidance will be facilitated by developing and making available training in the use of AI that is freely available to allow everyone the life-long possibility for learning about whatever interests them.

 (Policy statement on developers of AI being required to publish environmental impact of the use of their technology and commit to making it carbon neutral, water positive, etc)

DCMS 125 As with all technology there is the possibility for negative environmental impact, either foreseen or found during use. To avoid this, the developers of AI will be expected to provide reasonable measurements of environmental impact such as carbon footprint, water usage and others with the aim of reducing these to zero or positive impact as soon as possible.

(Policy statement on Intellectual Property Rights)

DCMS 126 AI platforms that can generate written text, music, images or realistic films will automatically be categorized as High-Risk and subject to regulation. This regulatory framework will make clear the way in which an individual can claim unfair use of their own creative output and curtail its continued use.

DCMS 127 The "training material" used for developing AI systems such as Large Language Models should be restricted to non-copyrighted internet content. However to avoid the potential for large corporations buying up significant portions of existing content, this requirement shall be balanced against the public need for the results of research and well-informed content to be used in training AI that is unbiased and up-to-date so it will be subject to independent review where conflicts of interests arise.

(space for policy statements on any other areas I've forgotten)

LONG-TERM POLICIES

(Policy statements on developing bodies to oversee AI development)

DCMS 128 As stated above we will foster international co-operation towards the establishment of a world-wide body similar to the International Atomic Energy Agency that could oversee and police AI development globally, and a similar body within the UK. Long term we will work towards the UK establishing itself as an important leader in bringing all AI policies world-wide in line.

(Policy statements on goal alignment and the need for policy to continually evolve to ensure it meets current needs.)

DCMS 129 Long term policy on an emergent technology and its largely unpredictable development cannot be drawn up with certainty, however the broad aim that should persist is that we guide AI so

that its goals are aligned with those that we consider best for humanity and the environment via a system of regulation, checks and guidance that can evolve as time progresses.

References:

1) https://www.safe.ai/statement-on-ai-risk
2) https://artificialintelligenceact.eu/assessment/
3) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4064091
4) https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/ai-health-and-care-award/ai-health-and-care-award-winners/