DEBATING RESPONSIBLE AI
The UK expert view

Ipsos

Contents

Contributors

Ofcom

AI Council

Palantir Technology

Office for Statistics Regulation

Leverhulme Centre for the Future of Intelligence

Connected by Data

Department of Science and Technology Studies, University College London

King's College London

Responsible AI UK

Leading experts in responsible AI from the above organisations took part in the interviews, agreeing to do so on an anonymous basis. The findings set out in this report are based on Ipsos analysis and may not represent the views of these individuals or the organisations they work for.

Introduction

This Ipsos UK paper takes a deep dive into one of the pressing technological and social issues of our time: responsible Artificial Intelligence (AI). The paper draws insights from ten distinguished experts in the field based in the UK, including eminent academics and leading regulators, each contributing their unique perspectives on this critical subject.

As a rapidly evolving technology, AI has begun permeating every facet of our lives, promising immense benefits for individuals, organisations, and societies. However, it also brings complex challenges and risks that emphasise the importance of developing and using AI responsibly. This paper describes the current debates around defining responsible AI, the many benefits and potentially significant drawbacks of AI, and the role regulation can play in enabling responsible AI.

It is our intention that this paper sparks thoughtful discussion around responsible AI – what it means, and how it can be applied in practice.

What does responsible AI mean?

Reflecting the complexity of AI technologies and differing views about its use, experts frame and describe responsible AI in different ways. AI has already been in use in consumer products for the best part of a decade, but the recent rapid emergence of generative AI has brought long-standing ethical questions to the forefront of public debate.

There is much discussion about whether and how AI can be developed and used *safely, ethically, or responsibly*. Sometimes these terms are used interchangeably, sometimes to different ends and purposes. Experts in the field of responsible AI recognise the divergent views on the meaning of these terms and how they relate to one another.

This means there is no overall consensus on what responsible AI means in theory or in practice. However, there are some key themes and areas of debate that emerge across the interviews, underlining different expert priorities and starting points.

❝ Industry has a responsibility because they are building this technology and they need to be thinking about how it's evolving, the principles and guidelines ❞

Responsible AI requires a broad, socio-technical approach

Many experts point to a need to look beyond the immediate risks that AI technology poses to considering properly the impact of the technology. They raise examples of the potential impact on society, places, jobs, skills, and the environment.

For experts, responsible AI therefore needs to take a 'socio-technical' approach that considers the wider consequences of AI development and use. This requires a broader understanding of risk and harm that goes beyond individual technologies and their application in specific contexts – to consider economic, legal, technical, political, and wider societal impacts.

*"[Thinking about economic questions gives us a wide vision of responsible AI], and I think this is why asking **economics questions** is quite **a powerful way to critique the frame of ethics** that's been allowed to dominate … the frame of ethics might not problematise the sorts of things that economists might be interested in, which would include things like **who is developing that technology,** and **what incentives are they following?"**

A participatory, inclusive, and context-specific definition

At the same time, experts also argue that society needs to agree what responsible AI looks like, and this often requires bottom-up or context-specific approaches. Without considering context carefully, defining responsible AI too prescriptively could result in an approach that lacks sensitivity to cultural norms or different priorities at different times across societies – this is the 'socio' element of the 'socio-technical' approach.

There is also concern that too rigid a framing might overlook uncertainties or unknowns when it comes to taking a responsible approach. For example, as one expert explains, tech leaders currently rely on the public to trust implicitly – based on promise alone – that tech companies are developing responsible AI. However, wider society is not taken along on the journey to understanding what risks are being managed and how.

*"When it comes to AI, most of the **time when powerful people talk about Responsible AI** what they mean is, "We are doing the responsible thing, don't worry we'll take care of it." I would say that's an instrumental deployment of the responsibility; it's **antidemocratic definition of responsibility** … There's a sort of **exclusive use of the idea of responsibility rather than an inclusive one** … the more inclusive way would be to say **we don't actually know all of the issues and all of the questions around AI."**

As a consequence, some suggest that a clear articulation of responsible AI requires democratisation and drawing from the evidence and expertise of a wide range

> **❝ When powerful people talk about Responsible AI what they mean is, 'We are doing the responsible thing, don't worry we'll take care of it' ❞**

of stakeholders. This would include AI researchers, developers and industry organisations, social scientists, philosophers, economists, ethicists, policymakers, industry leaders, community representatives, and diverse publics. These perspectives have the potential to offer a more nuanced and forward-looking understanding of responsible AI, touching upon theoretical, practical, and philosophical dimensions.

Experts feel that a collaborative approach is essential to determining responsible AI, and within this, different stakeholders have a different share of responsibility.

*"**Researchers** have a set of responsibilities which we are addressing and **in terms of introducing ethics** into all of the work. **Industry** has a responsibility because **they are building this technology** and they need to be thinking about how it's evolving, the principles and guidelines that have been set for them in terms of doing this, and the fact that they are out-pacing regulatory and policy making right now. And then finally, **government and policymakers** have a responsibility because **they are the final backstop**."*

*"Our approach is very much using our existing **regulators to police things** like the Online Safety Bill, and that is crucial – it is **crucial to AI going forward** – what can people produce, what are the limits of what people can do with AI."*

Broadly, experts define the roles of responsibility in terms of 'solution owners' (the developers of AI) and 'challenge owners' (the industries using AI).

*"The opportunities under this umbrella [of responsible AI] is that it is much more inclusive, and **it lends itself to a much more multi-disciplinary approach**. AI Ethics was already leading us towards being multi-disciplinary. When you start to think about responsible production of technology, you have to take into account experts, what I call **solution owners and challenge owners … these two people need to work together,** not just when the solution owner has built a thing and then tries to stick a square peg into a round hole, but right at the beginning."*

Beyond this, experts feel that through this collaborative process there is an onus on society to engage with AI and shape what responsible AI looks like. Experts from academia, research

" It is crucial to AI going forward – what can people produce, what are the limits of what people can do with AI "

organisations, and the technology industry need to help by closing the knowledge gap and empowering society to do so. The technology sector is also responsible for following guidelines and policies, ensuring the development and use of AI is suitable and transparent so that people and organisations have clarity about how their data is being collected and used.

"Society has an onus to learn and understand and take leadership of themselves in this digital age. We as experts have a responsibility to make sure that the skills available or imparted on them are fit-for-purpose. They don't have to know everything about everything – we have to do the hard work to make it easy for them to take that responsibility."

"You need to have some kind of explainability, possibly even replicability, to give assurance that what is done with the training data is appropriate and understandable by a human. The challenge with something that mimics a neural network is that it is so complicated. It is actually quite hard, even for a deep expert, to understand how a particular result has been arrived at."

Moving from principles to practice and political choices

Some experts highlight the importance of understanding the principles and ideals that underpin the development and deployment of AI systems. But they also recognise that it is just as important that these principles are put into practice and embedded in the organisational culture or consistency of decision-making practices.

More practical approaches to conceptualising responsible AI might start from the standpoint of considering individual privacy, risk, and harm. However, experts usually argue that this too should broaden out to consider wider societal and political impacts, for instance for workers and industries, and for equity. This might involve

considering whether, for instance, the design of AI technologies entrenches wealth, power, and opportunity where it already is situated, rather than distributing those more broadly.

Some experts went further, critiquing notions of ethical AI as too focused on the expert-based ethics discourse and too elitist and remote from the practical considerations of everyday citizens. Instead, they encourage a focus on the politics of responsible AI.

"The tech industry and ethics has been really problematic and we've seen it in a number of hugely well-funded centres for ethics research and a relative neglect of questions that I would regard as those of politics which, according to simplest definition of politics is about who gets what and how."

> **❝ The challenge with something that mimics a neural network is that it is so complicated ❞**

Responsible data use must underpin responsible AI development

The role of responsible data use is another priority for experts. Responsible AI is built on responsible data use and management across the data lifecycle – ensuring responsible data collection and use, robust model development, explainable and interpretable AI, and clear accountability mechanisms for continuous monitoring and auditing.

Experts also emphasise the need for clear and accessible guidelines and oversight to facilitate the responsible deployment of AI, addressing issues like algorithmic bias, data inequity and privacy, and the efforts to address potential differential consequences of AI use for different communities.

*"Did it have a **fair and proportionate data set to be trained on** and therefore **the outputs are relevant to my particular population** for my particular decision making or my particular action?"*

*"Is the data training **free from bias, appropriate, reliable,** and so on; and **do we have the consent of the data owners** (whoever they may be) for the training data to be used."*

AI needs clear purpose avoiding negative consequences

Some experts point to the importance of clarity about the goals of using AI technology itself as crucial for a proper understanding of responsible AI. This means ensuring AI is being developed and used to address some of the most pressing challenges faced by humanity. Examples given included AI-powered diagnostic tools which have enhanced the accuracy and speed of medical diagnoses, while machine learning algorithms enable more efficient resource allocation in disaster response efforts. Guarding against unintended negative consequences is important, as is taking action when they arise.

*"If you're building something, you should be saying, "OK, I'm going out to build it for this purpose or to solve this problem". So, it's **a very clearly defined problem statement** and that's what I'm going out to do. If I happen to find it does something else, then I have a **responsibility as a developer, a company, or institution** – whoever I might be."*

> **" If you're building something, you should be saying, 'OK, I'm going out to build it for this purpose or to solve this problem' "**

Tackling discrimination, bias, and creating an equitable society

Linked to this question of purpose, experts emphasise the importance of thinking beyond the effectiveness of AI in isolation, and even the immediate individual rights and obligations that the AI technology might generate – such as privacy, explainability and transparency.

Experts also suggest responsible AI encompassing the impact of data-driven and AI systems on different groups, societies, cultures, and ways of working. They emphasise a focus on the equitability and accountability of those systems to different groups, and on ensuring they enable and encourage a sense of agency, control and freedom amongst people, as well as considering how benefits and harms are distributed.

*"One thing you might be concerned about is **how does that system perform**, but also you might be concerned about **questions of accountability**: how do we know if **this thing makes a mistake** or misses something?"*

Clarity of regulation and hard lines agreed by consensus

Given this, experts emphasise the vital role for regulation in shaping the development and use of responsible AI. It serves as a critical safeguard, ensuring AI technologies are aligned with ethical, legal, and societal standards. The benefits include the potential for effective AI regulation to set clear guidelines for AI system design and deployment. Experts' suggestions for using regulation to support responsible AI is considered in the next section of this paper.

*"There **needs to be some hard lines** where you say to yourself, "We're not going do that." So, for example in healthcare, ... would you draw a hard line in trying to do gene alteration for changing the age of a cell so you can live longer? I know people are trying to do stuff in longevity, **but at what point do you draw the line** and say, "No, you're not allowed to do that"?"*

*"We now have technology that can instantaneously touch an enormous amount of the population and inadvertently affect them and **we don't have the right guardrails** in place for that."*

" There needs to be some hard lines where you say to yourself, 'We're not going do that' "

The opportunities and risks of AI

Experts are in agreement on the enormous potential of AI as a set of general-purpose technologies that can transform many areas of business, life, and society. But when it comes to ensuring the responsible development and use of AI, experts' attention is understandably focused on addressing the risks.

Indeed, on balance, the UK public think AI is more of a risk than an opportunity (*Ipsos UK KnowledgePanel*, September 2023). These risks relate to both the use of AI technology and the wider impacts discussed previously.

The primary focus for experts is therefore how to mitigate the risks and implement guardrails so that the many potential benefits can be realised without causing undue individual and societal harm.

*"AI technology is just like any technology: it can be used for dual purposes, **a double-edged sword**. It is the responsibility of all of us to **make sure that the guardrails are in place**, more so than other things. So, with the invention of a fork or a spoon, I could use that also as a weapon, for example, and that is a bit of technology creating a fork or a spoon, but we have **rules against hurting other people**."*

❝ AI technology is just like any technology: it can be used for dual purposes, a double-edged sword ❞

Defining the risks

Experts broadly describe the risks of AI in two categories – those that are known or emerging (e.g., algorithmic bias related to model training, the proliferation of misinformation via large language models, etc.), and those that are unknown or have yet to emerge (e.g., large-scale job losses because of AI automation, or the end of humanity due to an overreliance on AI technology).

This means accounting for a range of potential risks as broad and varied in scope as the potential applications and opportunities associated with AI. Experts acknowledge this breadth of risks as a key challenge in itself. Given the wide applicability and still nascent stage of development of many AI technologies, collating and understanding the shape and specific nature of these risks is a hurdle to overcome in the development of responsible AI.

*"Am I fearful of it? I think I just don't understand it – none of us do. It is the same as **standing on the edge of a black hole** – the theory and understanding we have of physics: gravity, friction, and all this kind of stuff falls away in a black hole, or **we just don't know which one of those rules stand.** I think we are faced with the same thing."*

*"There is a **really wide range [of risks]**. I sound pessimistic, but that's because the technology **can be used in so many different ways**."*

Assessing the risks

Digital technology is already embedded in many areas of people's lives, and this is the context for the rapid development and integration of AI in consumer-facing technologies. Experts recognise that the public – and many organisations outside the technology sector – understand little about AI, and therefore anxieties about the potential threats of AI are now prevalent in the zeitgeist. Indeed, the majority of the UK public say they only know a little about AI (*Ipsos UK KnowledgePanel*, September 2023).

In the public imagination, the emphasis is often on unknown risks, where the threats of AI are commonly understood in existential or dystopian terms. Here experts allude to the Hollywood 'mythologies' of AI created by the likes of Stanley Kubrick or James Cameron. Typically, the imagined risks take the shape of

> ❝ **Am I fearful of it? I think I just don't understand it – none of us do** ❞

a post-apocalyptic nightmare – one in which we submit to our computer overlords after they have achieved sentience, or otherwise a reality in which robots wreak generalised havoc on human civilisation. Similar views are explored in Ipsos' companion piece 'Very Human Reactions to AI'.

*"There are **real dangers** and of course there are **lots of ways in which these systems can be used for ill by people who tend to do so**. Those are very important risks. At the other end of the scale there are existential risk worries. I think **the one about the machine uprising gets disproportionate attention** because we are so familiar with it – AI has a mythology around it."*

Prioritising the risks

When it comes to responsible AI, the emphasis for experts is rather on the known risks. They prioritise addressing the problems associated with AI already impacting society, such as algorithmic bias or the related issue of equitable and transparent automated decision-making at scale. There are also the emerging risks around generative AI's potential to proliferate misinformation, impact critical thinking, or

exacerbate online harms – for example making hacking easier or crimes such as identity theft more convincing.

Given there are real harms already emerging which could impact society at scale, experts feel that in practice this should be the short-term priority for responsible AI. They typically see the potential existential risks as real but less immediate. This attitude is generally shared among the UK public, with the majority thinking it is unlikely that AI will lead to the extinction of the human race (*Ipsos UK KnowledgePanel*, September 2023).

Experts express concern that focusing only on these longer-term catastrophic risks could be at the detriment of the known risks and problems that matter now. They therefore believe priority should be given to addressing known risks that fall into economic, cultural, social, and political

domains, and not only or primarily the unknown existential risks.

*"I think **we should worry about a whole range of things**, so I don't want to suggest no one should be thinking about the loss of control terminator scenario – I think some people should be. But my worry is that we lose sight [of other issues]. When people think about AI policy, **are they thinking about the possibility of a million people losing their job in a relatively short timescale** and the immense social disruption that will cause."*
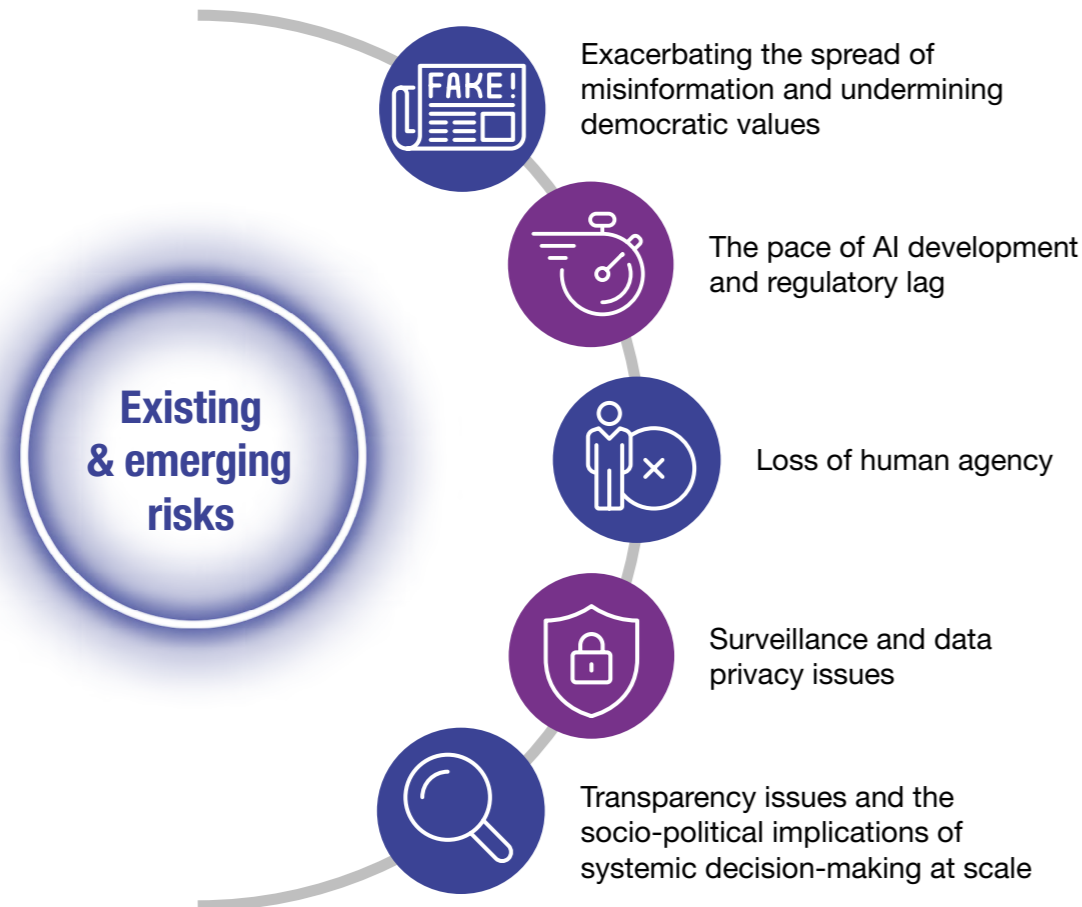
These types of risks are also not far from the public's current concerns. While most think businesses and healthcare patients will benefit from AI, the majority think that unskilled workers will lose out (*Ipsos UK KnowledgePanel*, September 2023).

When people think about AI policy, are they thinking about the possibility of a million people losing their job in a relatively short timescale

Types of existing and emerging risks

This means that among the known and emerging risks of AI, those that are top of mind for experts are not new, but rather issues familiar from other industrial and technological shifts, or those associated with existing algorithmic and data-driven technologies. The widespread use of AI means these have the potential to become more complex and have larger-scale impacts across more domains.

Existing & emerging risks

- Exacerbating the spread of misinformation and undermining democratic values
- The pace of AI development and regulatory lag
- Loss of human agency
- Surveillance and data privacy issues
- Transparency issues and the socio-political implications of systemic decision-making at scale

Following the rapid emergence of generative AI, there are considerable risks around large language models proliferating misinformation, with an impact on public discourse and democracy.

*"Then you've got the broader set of people who are affected by the results – the use of **the output of that generative AI**. So, that's where **issues such as misinformation, disinformation, deep fakery** that kind of stuff kind of comes into play."*

Experts agree that a key challenge is the potential risk posed by an AI regulatory lag, where the pace of AI development and use outstrips the capacity to create policy frameworks and appropriate regulation. So, even if the risks of AI are all known, we may not have the regulations in place that are necessary to address them.

*"The other thing which is **the more interesting challenge is the rapidity of how stuff moves in data and AI**. I think that's interesting because it calls into question **whether the kind of timelines that we have for legislative changes are**

enough and **whether we need different kinds of levels of policymaking** *that enable the system to respond faster to new challenges as they come up."*

The capacity of AI technologies to make decisions at scale is a key risk for experts as it undermines human control and agency. Broadly, there are two risks within this – first that the decision-maker loses control of the decision, and second, that those directly impacted by the decision lack agency, including understanding and challenging the decision.

*"In **automated decision-making**, you've got two relationships that you need to care about: **how is it affecting the people whose data is being collected**, and **how is it affecting people where decisions are being made** about them?"*

The rapid development and use of AI is also complicating the yet unresolved tensions around data privacy. Among experts, there a strong feeling that applications of data technologies including AI rely on techniques that are invasive.
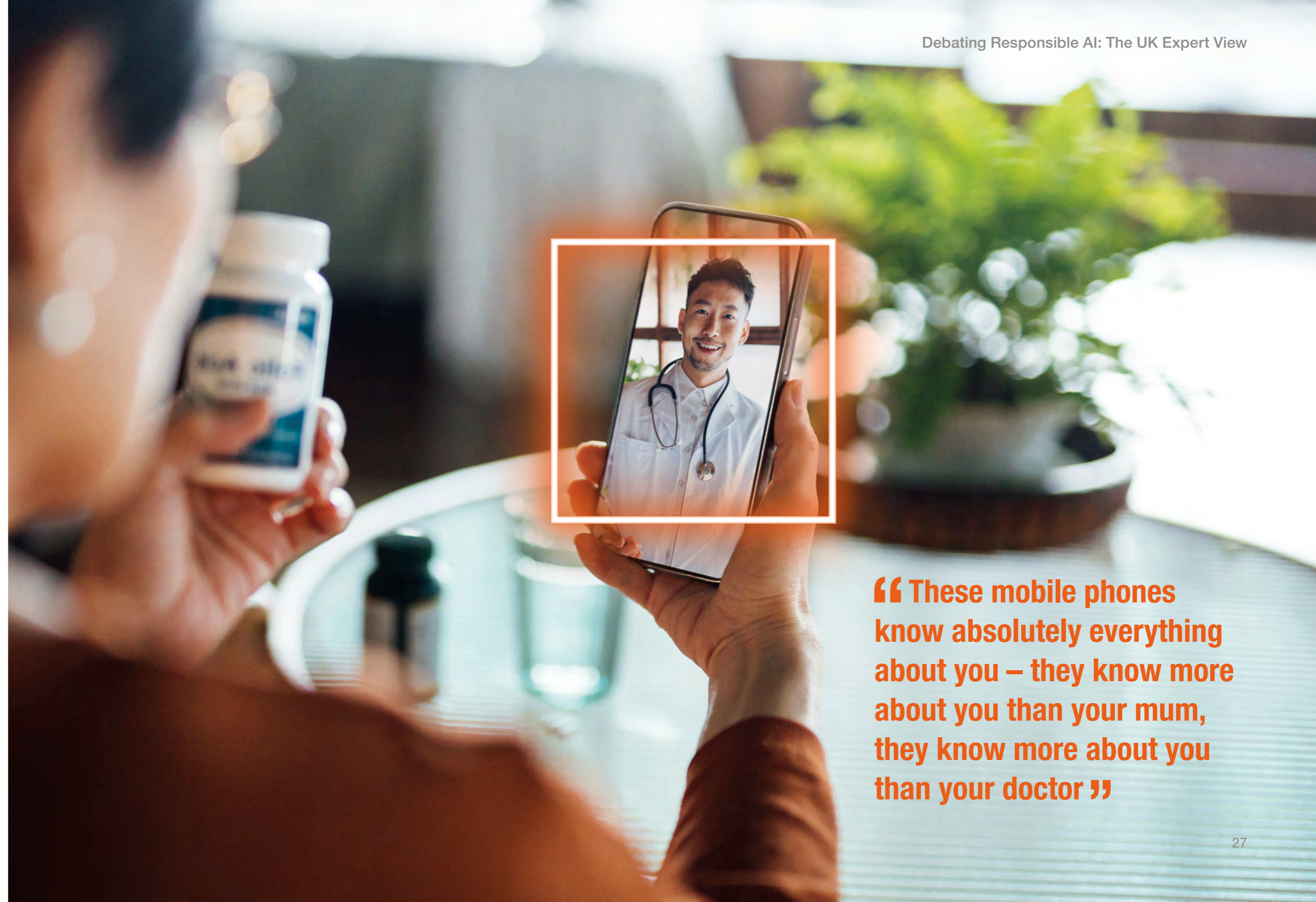
*"There is the whole surveillance side of things: right now, **these mobile phones know absolutely everything about you** – they know more about you than your mum, they know more about you than your doctor, **they know more about you than any other thing or person in your life."***

———

There is also a concern that if AI is relied upon to make decisions at scale, this could magnify the scope of the potential harms on the social and political landscape. AI technologies currently lack transparency, and therefore the decision-making process is largely unknown or opaque at best. It is therefore essential that technology companies provide clarity on how AI systems are being trained and applied.

*"[AI is] **the local lord making the decisions**. Bureaucracy is about making decisions consistent and efficient and one of the complaints people have about bureaucracy is that it is **often untransparent**, opaque, and if you set the parameters wrong then **making decisions efficiently at scale can be catastrophic."***

❝ These mobile phones know absolutely everything about you – they know more about you than your mum, they know more about you than your doctor ❞

While these known risks are experts' priority, there are still the unknown (or not fully known) potential risks to consider. These are not limited to the existential risks, and they have the scope for enormous disruption to society, as in the case of mass job losses.

*"It isn't necessarily women or ethnic minorities who have been marginalised systematically and whose marginalisation is being replicated by the algorithm. It **might be groups like working class white men who suddenly find themselves out of work at scale and consequently disruption.** So, those kinds of injustices – someone losing their job, whatever it might be – at scale can lead to **tipping points that would have catastrophic consequences."***

There are differing views among experts on how likely existential risks are to be a real issue. However, one common characteristic of the risks associated with AI – whether real or imagined – is the potential for human-designed technologies to have unintended consequences. This is where the mythologies of AI meet the practical realities. Within the current discourse around AI there is a very real fear that science can go wrong, as has long-since been mythologised in science fiction classics like Mary Shelly's *Frankenstein* – the fundamental concern is that the creation can go out of the bounds for which its creator designed it.

*"The **scarier stuff is at that top end** – then **it very quickly becomes science fiction**, or rather,*

*the best examples have already been written about by futurists and science fiction writers. **If we come back to reality, we have right now tech that exists** with the ability to create and replicate itself in thousands and thousands of bots that send out misinformation."*
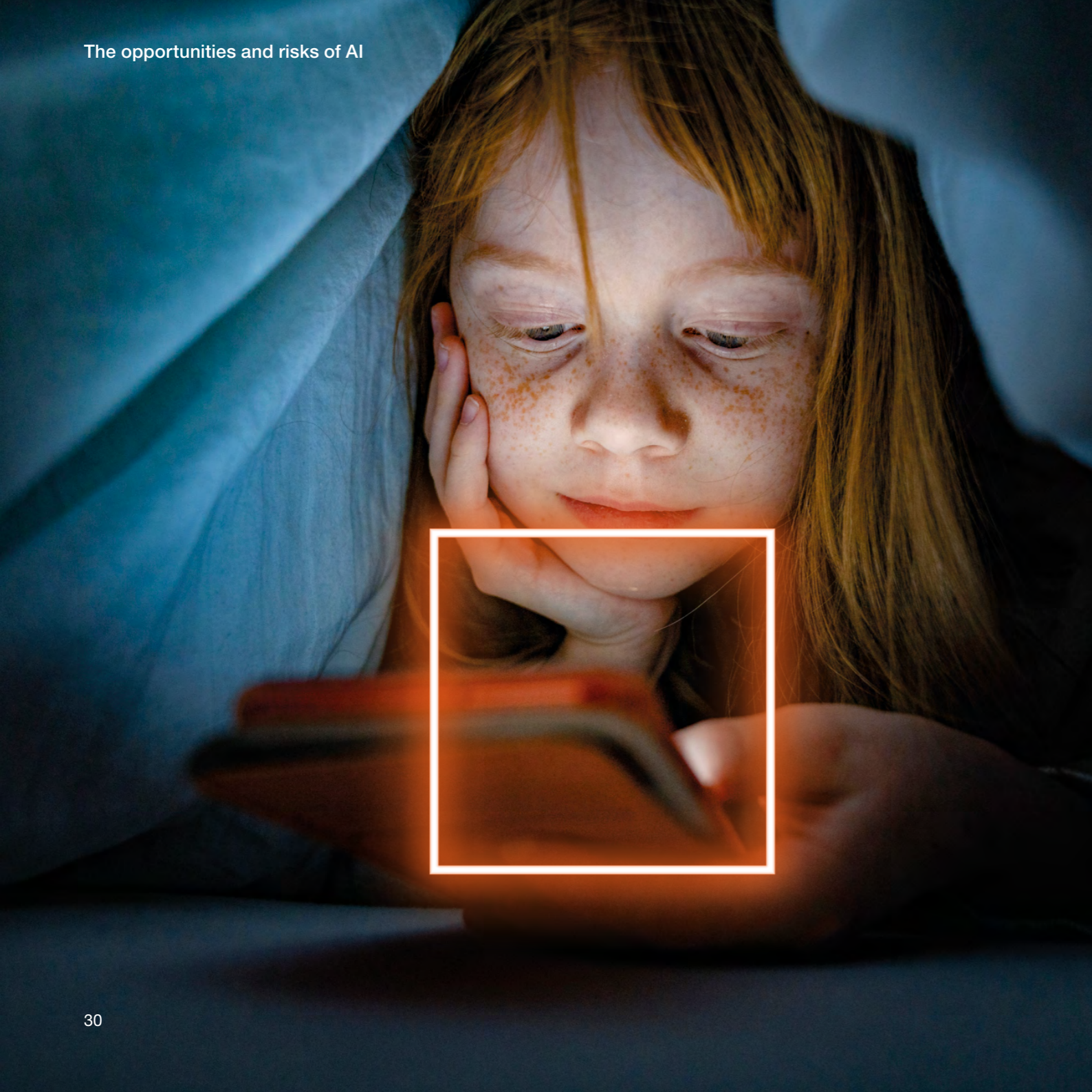
*"I think we need to be careful – there are **a lot of unintended consequences for these technologies** and there are a lot of ways to use technology (that is meant to be used well) in a negative way."*

In terms of assessing the more immediate potential harms of AI, experts feel that historic precedent in technological innovation (and the past impact of new technologies on society) provides a useful roadmap of the risks that should require regulatory prioritisation. While the nature and scale of the technology is new, societies have faced many similar challenges before. For example, with the arrival of AI, there are echoes of the emergence of radio in the early 20th century, and the concurrent spread of political disinformation and rise of the political right.

*"I'm like "Have you read any 20th century history? Do you know what happened **with the rise of***

> **❝ The scarier stuff is at that top end – then it very quickly becomes science fiction, or rather, the best examples have already been written about by futurists and science fiction writers ❞**

the radio and how that allowed the spread of Fascism. Do you know what happened with the rise of television and how every coup during the cold war started with claiming the TV stations". Misinformation is not new, **control of the flow of information is not new**. *If anything, it's become harder to take control [of the flow of information], but that doesn't mean we should be complacent."*

Realising the benefits of AI

Given the enormous potential of AI, stakeholders believe it is in everyone's interests to ensure that the technology is developed and used responsibly. There is wide recognition of its scope to be applied across sectors, from businesses to the public sector, and for it to address the current and future issues society is facing such as racial inequalities and preventing harmful content.

"My mind immediately goes to the justice system and the way that the justice system works. So, we know – I mean this is especially true in the States but it's also true here – that there is systemic racism in the justice system right from policing through to sentencing and parole hearings and everything … you could create systems that correct for the systemic racism and that give a fairer outcome that for example doesn't over police the black neighbourhood."

"We are looking at the use of AI for estimation technology – so, estimating people's ages to prevent them from accessing adult material. We are using AI for certain types of quantum moderation to pick up through pattern detection what posts on a social media platform might be harmful, derogatory, containing hate speech, containing terror content, CSEA[1] content, etc. All of that has got a lot of great promise."

---

1   Child Sexual Exploitation and Abuse

You could create systems that correct for the systemic racism and that give a fairer outcome

Regulating for responsible AI

Experts emphasise the importance of regulation for achieving responsible AI, and most of public think government in the UK and internationally (along with tech and social media companies) are currently doing too little to regulate AI (*Ipsos UK KnowledgePanel*, September 2023). Experts think effective regulation needs to describe what responsible AI means in principle and in practice in different contexts. Regulation will also need to develop enforceable rules and explain the required actions – and consequences – when things go wrong. Given this, experts see regulation as taking the lead in defining, assessing, and mitigating the risks associated with AI technologies.

As discussed earlier, experts consider regulation as crucial for managing the risks of AI while making the most of the opportunities of these technologies. Importantly, getting regulation right can reassure the public that AI is being

developed and used appropriately, helping to build public understanding and confidence. Regulatory frameworks also provide clarity for organisations as they develop and seek to apply AI in different contexts.

"You absolutely need regulation, not to stamp out innovation, but to allow innovation to happen in an organised, responsible, and safe manner for society."

"Most stakeholders who are developing or adopting AI, to be honest, will only do things they are compelled to do by the legal frameworks in the sectors that they operate in. If there is some kind of unintentional harm upon consumers, they're more likely to say, "There's regulation missing, so you need to regulate."

However, developing regulatory frameworks that can be applied effectively across the huge range of use cases for AI is a considerable challenge. While there are emerging ideas for the types of approaches that might help, experts generally describe this as an area where more work is needed. Instead, they point to a series of considerations for developing appropriate regulatory frameworks for responsible AI. These link to many of the issues raised previously in this paper.

1. Regulation based on principles

Experts discuss the importance of agreeing overall principles for governing how AI can be developed and used responsibly. They usually see setting these principles as a role for government, but also point to the importance of sector and public involvement in developing priorities and considering the trade-offs and red lines. As described above, experts think framing regulation in terms of safety alone is not enough – the wider societal impacts of AI should also be considered.

"You might want to ask – and this might inform your regulatory framework – "What effect do [applications of AI] have on markets … what effect do they have on political systems?" So, you might want to look at risk in a much, much broader way."

Experts recognise that developing general principles for responsible AI is complex. One obvious challenge is the wide-spread lack of understanding of AI: among the public, businesses, and policy makers. Deciding on these principles will require an informed debate about the capabilities of AI technologies, the potential impacts on society, and an understanding of the risks and benefits for a range of different use cases.

> **❝ You absolutely need regulation, not to stamp out innovation, but to allow innovation to happen in an organised, responsible, and safe manner for society ❞**
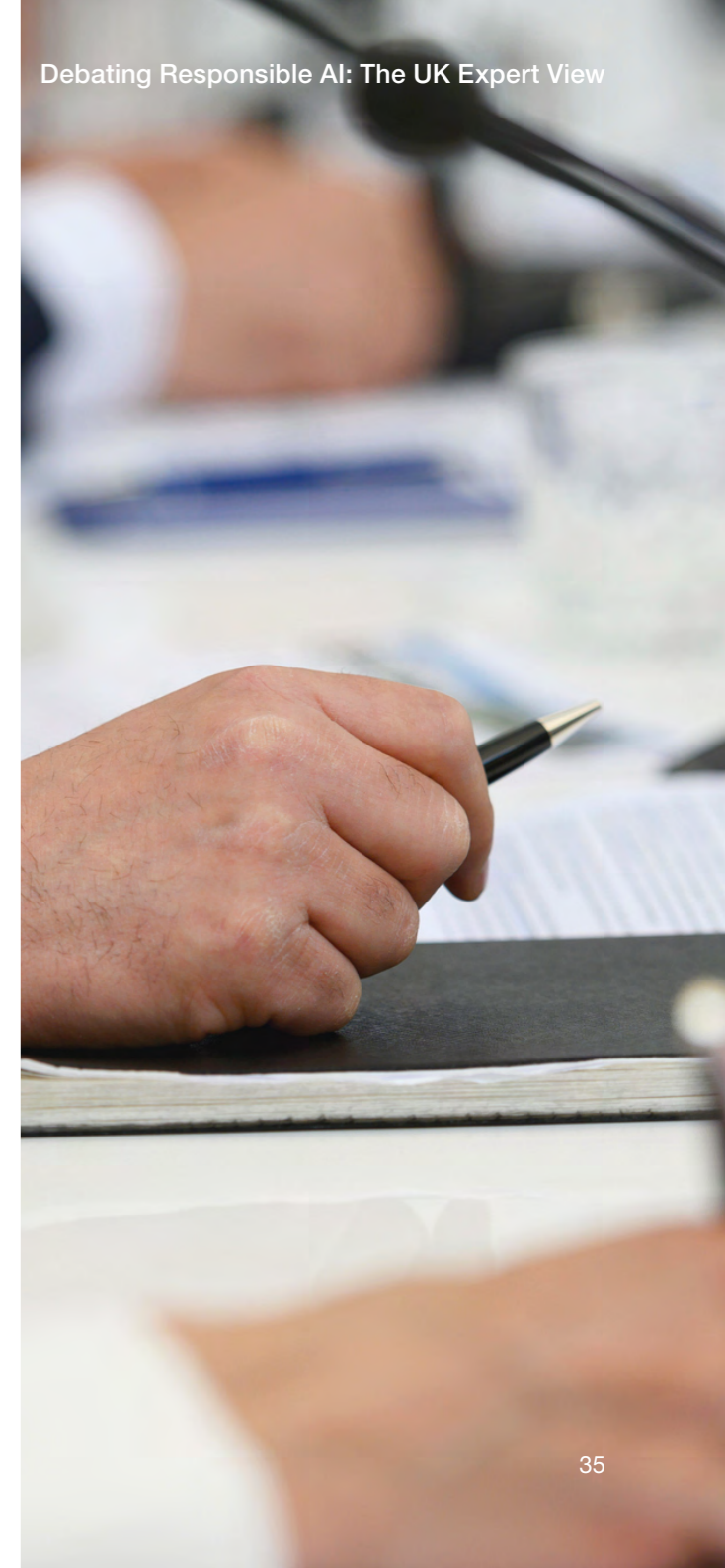
2. Regulation tailored to specific contexts

While vital, experts feel that agreeing principles for the responsible use of AI is not enough on its own. The wide applicability of AI technologies in complex systems and the range of potential individual and societal impacts means a 'bottom-up' approach is also necessary, as discussed previously. Experts stress that AI regulation should be adapted to specific use contexts and cannot simply be based on general rules grounded in the characteristics of a particular AI technology.

*"I think that the **crosscutting regulations** around data and AI systems struggle because they **can't imagine every context.**"*

This is broadly reflected in the UK's sector-based approach, and there is some support for regulating AI in this way. There are also concerns about what a sector-based approach might mean for applications of AI that do not fit with an existing regulator's remit. The different expertise and capacity of sector regulators is also a challenge.

*"A **sectorial approach** in which different regulators are approaching and evaluating a system within their domain … seems like a good idea because there's such a wide remit of different technologies. But it does raise **a huge issue in that not all regulators are the same** – not all regulators have the same capacity."*

Tailoring can also take the form of a proportionate, values or risk-based approach, such as in the EU AI Act. Some experts think this approach has merits and could form part of the tailoring applied in the UK context, and starting point as to what a socio-technical approach to the risks of AI means in practice.

*"I guess the challenge is making sure that the response and the safeguards that are in place for those different systems is proportionate. **Google Maps is important but if it goes wrong** and it sends you to the wrong address **it's not as problematic as if a cancer diagnosis system gets your diagnosis wrong.**"*

However, there are also concerns that the risk-based approach adopted by the EU AI Act may be too prescriptive and limited in scope,

particularly given the challenge that many of the risks are yet unknown.

*"The **challenge with the EU AI Act** is that, as far as I remember, **they've dictated what those risk levels** are from the centre from the beginning. I'm sure they can change it, but they've said from the get-go that XYZ is high risk and ABC is low risk – **I don't know how well that will hold up in reality** … whereas the UK's model is far more flexible."*

*"I think **the right kind of framework will be values-based like the EU one.** It will have real measures in place that require the responsibilities of the technology companies like they are required to use personal data in responsible ways since GDPR and the introduction **of legislations ensuring there is real accountability.**"*

Experts argue that future-proofing regulation is another important reason for ensuring frameworks can be adapted to different contexts. The rapid pace of technological innovation is seen as a more general problem for regulators, but one that is particularly acute for AI. The underlying technology is constantly developing,

and use cases are being worked on in every area of business and society. This means regulation will have to be flexible enough to at least evolve in parallel – or ideally lead responsible AI development.

"It is no good just regulating for the current generation of things, **we have to think about what is going to come down the pipe** *and be prepared for that."*

As well as these longer-term concerns, experts also want to see regulation catch up quickly with the risks and harms already associated with AI. For example, developing agreed approaches to assessing and dealing with bias in decision-making is a priority. This is particularly important before AI becomes more embedded in public sector organisations because their decisions can have substantial consequences for individuals. One idea several experts mention is the use of Impact Assessments when introducing new AI technologies.

> **" I guess the challenge is making sure that the response and the safeguards that are in place for those different systems is proportionate "**

3. Regulation providing legal clarity, accountability, and redress

Experts highlight that while principles and tailored guidance are crucial, a clear legal framework is also essential for achieving responsible AI. This framework should give clarity to individuals and organisations using AI or being affected by others using AI. Without this clarity about important issues such as transparency, accountability and redress, there is a concern that some organisations would introduce new technologies without the right safeguards in place.

*"Whilst **I love talking about ethics**, because that informs the debate, I think when we are coming to regulation you can't talk about regulating ethics – is a company ethical or not – **we need much harder measures** than that ... in terms of how we set up, we need to start talking about openness and transparency: when some organisation is using AI they need to declare that they are and how they are using it, and what they are doing with it. **What they are doing has to be tested and evaluated in some way,** which is not straightforward."*

*"At the moment it's still in the "Is it right, is it wrong, where do we as a society draw the line?" [stage], but **I hope it will then come to the point where [we say], "This is the line,** and if you cross this line, this is the legal implication of it."*

Experts agree that regulation should set out what should happen when things go wrong with AI technology. However, knowing when things have gone wrong is not always straightforward. The impact may not be obvious to individuals or even organisations using AI. Regulatory frameworks will therefore need to consider how best to monitor the impact of AI in specific contexts, and where responsibility for doing so sits. Linked to this, some experts argue that talking to groups affected by AI about the impact on them – positive and negative – will be essential for responsible AI.

Much of the discussion about AI regulation has focused on the technology companies developing the underlying foundation models. Experts point out that there are further challenges for other non-technology organisations that want to use AI models. The organisations will need to be confident that the underlying technology meets agreed standards for responsible AI but will not have the expertise to assess this themselves.

> **"** We need to start talking about openness and transparency: when some organisation is using AI they need to declare that they are and how they are using it **"**

This means that assuring AI technologies will be another important aspect of responsible AI.

As part of this, experts suggest that Impact Assessments and audits are likely to be useful tools for compliance, as they are in other sectors. These are relevant for government before introducing new policies or services, and for technology companies while developing AI. However, the huge range of uses of AI makes this challenging, and it is still not immediately clear how we should address them.

"Who gets what and how, who benefits, who decides – those are the gold questions of politics. When you're assessing any sort of policy initiative you would be asking these, you would **do some sort of benefit assessment on where that policy's impacts are going to land.** *I think it's absolutely vital to ask questions – the difficulty is that our answers are going to be chronically uncertain."*

"Concretely, I really think that having a big focus on impact assessment – what kinds of impacts need to be assessed, who needs to be involved in that impact assessment and how are the results of those **impact assessments** *published and*

made available – [we need to be focusing on that]. **If there was proper work on that then we would get a long way because it would force a process** *that would get to improved outcomes."*

4. Regulation that works globally

How to develop international AI regulation has been a consistent theme in the debate about responsible AI. The global reach of AI applications means many experts would like to see standards agreed across countries. Complying with different regulatory regimes will cause significant problems for organisations operating across national borders.

Given the complexity of developing international standards, experts think a comprehensive global approach to AI regulation is not immediately achievable in the short term – partly because of the challenges and difficult trade-offs involved. They also note that differences are already emerging between approaches in the EU, US and China.

Experts also have different perspectives about whether a global regulatory framework is possible

or even desirable in the longer term. Some want an international agreement to ensure a basic level of compliance, similar to other industries such as aviation. However, reflecting different national priorities and cultures is also important.

"The regulation can sit at a country level, that's fine, but **there needs to be international coordination and that body needs to have teeth.** *So, they need to be able to take action if they need to shut stuff down until further things happen."*

The global influence of the regulatory approach in the EU AI Act is an area of interest for experts. They generally welcome the EU's overall approach of grouping technologies based on risk and tailoring regulatory responses accordingly. However, there are some concerns that a single, centralised approach to determining technology risk levels may not work in practice, given the novel ways these technologies might be applied.

"[The EU] are really **struggling to grapple with how to deal with general purpose technologies** *like foundation models, which are use case agnostic – the use case comes up more in the application stage."*

Conclusions

Experts agree that developing and using AI responsibly is a unique and difficult problem. They argue responsible AI needs both a broad approach that considers the socio-technical impact of these technologies, but also an approach that is flexible enough to be inclusive and context-specific. This will ensure the nuances of different uses of AI are fully recognised and addressed.

This means responsible AI will need to move beyond principles to provide a framework for practical and political choices. For experts, responsible AI also involves clarity about why these technologies are being used, how this will be done with the right safeguards in place, while being conscious of the potential for unintended and harmful consequences.

Experts also agree about the huge potential of AI to transform the way we live, run businesses, and organise societies. But there are risks: those that are already known to have a negative impact, and those that are more theoretical or unknown.

The experts we spoke to recognise the importance of responding to the existential risks, but also emphasise taking steps to address known risks. These include the pace of AI development, the loss of human agency, exacerbating misinformation, transparency in decision-making and ensuring data privacy. Experts argue that addressing bias and fairness at every stage of the development and use of AI is a key priority for responsible AI.

While regulatory approaches are far from straightforward, they are seen as essential for responsible AI. For experts, agreeing shared principles across stakeholders, and building frameworks that reflect these principles in specific contexts will be necessary for effective regulation. Regulation that helps achieve responsible AI will also need teeth: providing legal clarity, with red lines and redress mechanisms obvious to all those involved. Regulation needs to think ahead to future developments in AI technologies too, and to recognise the global context for responsible AI.

These expert views give a sense of the challenges ahead in developing responsible AI. However, they further contend that using AI technology responsibly is the only way to ensure we fully reap the many potential benefits for people, businesses, and governments in the UK and around the world.

❝ At the moment it's still in the, 'Is it right, is it wrong, where do we as a society draw the line?' [stage], but I hope it will then come to the point where [we say], 'This is the line, and if you cross this line, this is the legal implication of it' ❞

Authors

Daniel Cameron
Research Director,
Public Affairs
Daniel.Cameron@ipsos.com

Chantal d'Offay
Research Manager,
Corporate Reputation
Chantal.Doffay@ipsos.com

Reema Patel
Head of Deliberative
Engagement, Public Affairs
Reema.Patel@ipsos.com

Rachel Phillips
Research Director,
Corporate Reputation
Rachel.Phillips@ipsos.com

About Ipsos

In our world of rapid change, the need for reliable information to make confident decisions has never been greater.

At Ipsos we believe our clients need more than a data supplier, they need a partner who can produce accurate and relevant information and turn it into actionable truth.

This is why our passionately curious experts not only provide the most precise measurement, but shape it to provide a True Understanding of Society, Markets and People.

To do this we use the best of science, technology and know-how and apply the principles of security, simplicity, speed and substance to everything we do.

So that our clients can act faster, smarter and bolder. Ultimately, success comes down to a simple truth: You act better when you are sure.

Contact us