Trust, Safety and the Digital Economy

The Commercial Value of Healthy Online Communities

Perspective Economics

Careful Industries

Ipsos

Contents

Sophie Wilson, Michael Clemence, Jamie Douglas, Daany Ajaib and Kyle Morris, **Ipsos**

Sam Donaldson, **Perspective Economics**

Rachel Coldicutt, **Careful Industries**

Acknowledgements

The authors would like to thank all the workshop and interview participants for their time, contributions and insights, and Prof Mary Aiken for her advice and support. This report was funded by DCMS.

Executive summary

Digital life is here to stay.

The differences between life online and offline were blurred even before the pandemic. Now the growth of the metaverse and the spread of digital currencies are just two of many new and emerging uses of technology that may further dissolve those distinctions in the years to come.

As the pros and cons of digital tools and services are better understood, the building blocks of a better internet are becoming more apparent. This study shines a light on one part of that ecosystem: the emerging trust and safety sector in the UK.

Our findings show that a growing number of digital businesses are realising the benefits of building and nurturing trustworthy digital environments. These benefits are not just good for digital businesses, but are good for the whole of society.

Specifically, organisations we spoke with recognised the value of trust and safety for:

- growing trust in their brands
- cultivating more engaged users, and higher levels of both staff and customer retention
- anticipating new and emerging regulatory measures.

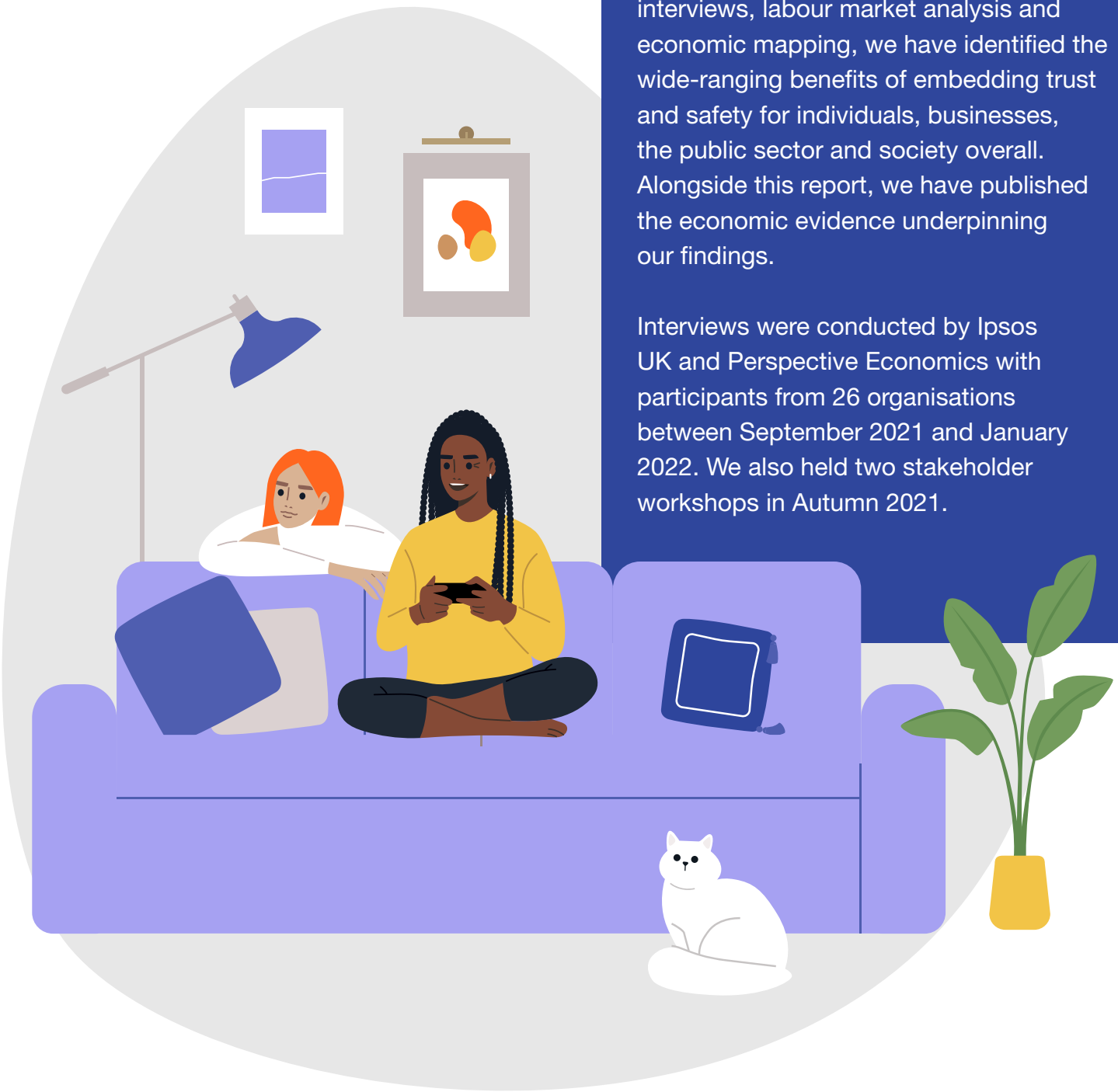
In turn, this has led to the rapid growth of a new specialist profession, with the number of advertised trust and safety roles across the UK economy almost doubling in 2021. The UK has a particular strength in the research and development of technologies which help companies spot harmful content, with the safety tech sector becoming one of the fastest-growing areas of the UK digital economy.

Over and above this, the pervasive nature of the digital world means that the advantages of trust and safety also extend to the broader economy.

Our research has identified a wide range of sectors, including infrastructure, media, retail, services and gaming, for which a greater investment in online trust and safety could provide dividends. A healthier, more sustainable and resilient online environment can provide **a key driver for economic growth.**

A healthier digital world also has benefits for our offline world in curbing the spread of hateful content, improving individuals' safety, and increasing civic participation.

However, our research indicates that the trust and safety sector is still in an early growth stage. There are opportunities to be had in nurturing the area as it matures, so that it grows responsibly, ethically and sustainably.



How this research was carried out

DCMS commissioned Ipsos UK, Perspective Economics and Careful Industries to explore the role of trust and safety in the UK digital economy. Through a literature review, stakeholder interviews, labour market analysis and economic mapping, we have identified the wide-ranging benefits of embedding trust and safety for individuals, businesses, the public sector and society overall. Alongside this report, we have published the economic evidence underpinning our findings.

Interviews were conducted by Ipsos UK and Perspective Economics with participants from 26 organisations between September 2021 and January 2022. We also held two stakeholder workshops in Autumn 2021.

Why trust and safety matters

Trust and safety is generally seen to cover the range of actions taken by companies to protect their online services against harmful content or behaviour, and to build healthier, more resilient online communities. The specific harms that trust and safety functions protect against include child sexual exploitation, the sharing of non-consensual intimate imagery, promotion of suicide or self-harm, hate speech and harassment, misinformation and disinformation, and spam and fraud.[1]

Spectrum Labs, a safety tech provider, defines trust and safety as follows:

"Trust and safety is the set of business practices whereby an online platform reduces the risk that users will be exposed to harm, fraud, or other behaviours that are outside community guidelines. This is becoming an increasingly important function at online platforms as they look to protect their users while improving customer acquisition, engagement, and retention."

Everything, all of the time

The role of online trust and safety has become increasingly prominent globally as digital technologies are fully established as a part of everyday life. Ofcom found that over 90% of UK adults use the internet, spending an average of three hours and 37 minutes a day online on computers, tablets and smartphones.

Ensuring social norms are upheld online is one of the biggest challenges being tackled by digital businesses, in particular among those who are seeking to cultivate online communities. The scale of this challenge can feel overwhelming, and the volume of user-generated content posted online is ever-growing. There is content that lives online forever, and content that disappears for the user almost straightaway. There is content that is shared publicly, content that appears only in private groups, and content that moves between platforms.

1. Cryst et al., 2021, p.1

"There is no definition of 'done'. As new communities come online with whole new ways to create content, these problems are only going to get harder and more complex. Twenty years ago this industry was only seen as chat filtering. Now, as a society, we know these challenges are much more complex and there are so many other things that you can do to cause financial ruin or harm to people."

Safety tech organisation

When designing services, creating acceptable usage policies and reviewing content, organisations need to carefully consider a complex range of issues. This includes child safety, freedom of expression, privacy, and the right of user transparency around any decisions made. This is a significant new shift. In response, more organisations are developing 'trust and safety' functions to support user health and wellbeing. The precise remit of this function varies, but there is a core of activity focused around:

- automation to support human moderators by detecting potentially harmful or illegal content at scale
- human moderators reviewing content and behaviour
- guidelines, policies and protocols to guide the work of these moderators and to protect users
- business governance and culture as levers for the creation of healthier online environments

The potential of automated solutions

Automated detection of illegal content can involve companies using content-matching or artificial intelligence-driven tools to analyse content at scale before detecting and flagging suspicious activity. This can be for the purposes of detecting illegal content on a platform, such as child sexual abuse imagery, or preventing illegal behaviour such as child grooming, violent threats and hate speech. Companies can also use automated solutions to help them monitor whether user-generated content is consistent with site terms and conditions.

Few companies have the resources to develop these tools in-house, as they typically require large and complex datasets against which to train and test artificial intelligence solutions. As a consequence, recent years have seen the identification and growth of an increasingly mature market in online safety technologies described as 'safety tech' - technology-driven products and services that help companies deliver safer online environments for their users. These technologies can also help to protect

moderators' wellbeing by reducing the amount of extreme content they are required to review.

Analysis led by the UK government in 2021 suggested that the safety tech sector was one of the fastest-growing segments of the UK digital economy, with average yearly revenue growth rates of more than 35%. Recent research by US firm Paladin has also identified a billion-dollar US safety tech market. Safety tech providers supply some of the world's biggest companies and online brands including the LEGO Group, Electronic Arts, Roblox, Riot, The Meet Group and Match, and football clubs such as Liverpool and Chelsea.

However, the use of safety tech is not always straightforward. For example, there is evidence that language models used in AI may amplify biases by associating specific words with negative or positive sentiments or stereotypes. Even where it is possible for malicious users to be detected and blocked on a site-by-site basis, there appears to be little information-sharing of knowledge between platforms. In this way, safety tech on its own is not a silver bullet. Ongoing discussions and iteration are required to fulfil its potential to protect users, coordinate responses and limit unintended consequences, including on freedom of expression and platform liability protections.

"It is a really rapidly evolving threat space … we're talking about people who are actively looking to subvert systems. We need to be alert to how these harms manifest over time and keep across those threats in the landscape."

Safety tech organisation

Empowering users

Technological solutions also have the potential to offer users greater control over their online environments. One example is the customised filters and controls offered by some Internet Service Providers (ISPs). Another is the emergence of apps such as Block Party which allow users to specify the types of contact they do, and do not, wish to receive from social networking platforms. These tools can play a useful role in helping users cope with the symptoms of online toxicity, even if they may not address the root cause. Online safety organisations frequently stress the need for technologies to be combined with media literacy training to help raise awareness of tools among more vulnerable, or less experienced, users, supporting them to navigate online environments.

"In order for people to make the most of life in the digital world, they need to feel confident in tech … People are better able to take advantage of our services if they feel more confident online."

Digital infrastructure organisation

Transparency

Several large platforms publish transparency reports that set out in some detail the actions being taken to enforce trust and safety on their platforms. These typically include actions taken in response to law enforcement requests, and to detect and address illegal content.

Although progress is being made by individual companies, the lack of established and consistent industry-wide accepted measures of user safety on digital platforms can make it difficult to truly understand, evaluate and measure the extent of harm taking place online. Interviewees for this report felt that greater granularity of how content is identified, processed and considered would be a welcome step in further understanding how decisions are made, and the variations across platforms and approaches.

Giving [the total number of harms reported] might seem great, and it is worthy, but you need to know if people trust your systems. Because if they don't trust your system, then people will just stop reporting, so you might in fact be doing far worse… There's a need for a lot more work in understanding what are the different types of behaviours that as a product we need to care about."

Gaming and entertainment organisation

These technical, ethical, and contextual challenges have been key drivers for the professionalisation and growth of the trust and safety community, particularly in the US with the Trust and Safety Professional Association, and the UK through the Online Safety Tech Industry Association. Both communities can explore relevant tensions, and new approaches can be developed, tested, and shared.

"No digital product or service we consume today wouldn't benefit from the application of safety technologies and safety driven approaches. We're at the start of an explosion in the sector."

Stakeholder organisation

Businesses are increasingly expected to step up

These debates are not taking place in a vacuum. Across the globe, citizens are demanding that companies and governments step up to secure their safety online. For example, research for Ofcom and the Information Commissioner's Office in 2020 found eight in ten UK citizens agreed that the most used websites and social media networks should do more to keep people safe, while just over half trusted those platforms to remove harmful content.

Business leaders themselves are increasingly recognising that the public is choosing to engage with companies who reflect their own ethical values. This reflects broader trends that we have seen in ethical consumerism, sustainability and diversity and inclusion. For example, in 2021 60% of the British public report buying from brands that reflect their personal values, up from just over 40% in 2013.[2]

According to recent research from safety tech company Crisp, more than 80% of CEOs felt that ignoring societal concerns could have negative material or financial consequences for business, with three quarters identifying that their company's values as a brand and employer will be tested by these risks.

2. Ipsos Global Trends 2021

Investing in trust and safety: The benefits to business

More people are demanding better and safer online experiences. This brings potential consequences for brands not delivering on user trust and safety in terms of both reputation and customer retention.

In our research, we spoke with stakeholders from 26 organisations from across market sectors and the safety tech industry, as well as public sector organisations, to understand their motivations for investing in trust and safety. The findings illustrate how organisations believe that protecting users is not just the right thing to do, but something which also translates into tangible business benefits. This is driving more businesses to invest in trust and safety upfront, as a way of both gaining competitive advantage and anticipating stronger global regulation.

Creating healthy, positive online spaces to retain users

Fundamentally, trust and safety is about creating positive spaces for users. Stakeholders emphasised how perceptions and experiences of positive online spaces can encourage individuals to return to platforms and spend more time there.

This creates positive associations with brands, building a longer-term relationship with users and creating opportunities to learn more about their interests and expectations for products and services. This has the potential to grow a company's user base, value and competitiveness, as well as to enhance product quality and consumer choice.

These factors cut across sectors, with platforms in gaming and entertainment, social media, dating apps and retail highlighting the importance of keeping users engaged on their sites.

> **"This goes back to 'trust and safety' as a discipline. We set the rules for the kind of communities we want to create and the kind of spaces we want for our users. That goes beyond what the law requires because we want that trust but we also want our customers to want to use our services without them being environments where they feel unsafe."**
>
> Digital infrastructure organisation

In the gaming world, efforts are being made to tackle toxicity and proactively encourage positive online behaviours. This is particularly the case for companies whose products appeal to younger audiences or explicitly cultivate children and young people as their user base, which may expose companies to expectations related to age assurance, age-appropriate design and moderation.

The range of potential responses are broad, and include providing policies, advice and guidance for human moderators; ensuring users have clear behaviour guidelines and reporting systems; and the use of moderation technologies. For example, Roblox deploys a range of safety technologies, including chat filters, customisable parental controls, avatar clothing detection and a reporting system. They also make use of a Trust and Safety Advisory Board to ensure that decision-making is informed by good practice as well as the experiences of parents and children.

"Ethically, it will be interesting to see if people move away from certain brands. I think they will assert themselves more. Gamers and those in the gaming industry are really loyal until they're not."

Gaming and entertainment organisation

However, implementing trust and safety approaches can come with trade-offs, and in some cases can lead to the deliberate introduction of friction into the user experience. For example, the LEGO® Life app has implemented Verified Parental Consent and digital citizenship materials aimed at children. The LEGO Group has also adopted pre-moderation approaches to prevent sharing of personal data and to restrict inappropriate user-generated content. Organisations report greater synergies between safety measures and engaging user experiences if considerations relating to user safety have been factored in throughout the design process through a 'safety by design' approach, rather than bolted on at the end.

"Parents would trust that our digital experiences are safe but if kids find it too disruptive, they'll move elsewhere."

Gaming and entertainment organisation

Online experiences drive offline engagement

User retention is not limited to online experiences. Stakeholders described how online experiences can influence how users feel about a brand, platform or service in general. Many organisations feel a duty to

protect their users or stakeholders offline and therefore implement trust and safety solutions online. In doing so, organisations are better able to respond to incidents of abuse in real time, stem the sources of hateful content and engage law enforcement where appropriate.

For example, some football clubs proactively address online harms and consider offline consequences such as stadium bans and legal action. Chelsea Football Club adopted Crisp's technology to identify, report and remove hateful and discriminatory posts and offers players support to identify and report abusive comments posted to their digital channels. Similarly, Signify AI used their Threat Matrix platform in collaboration with the Professional Footballers' Association to identify and report abusive accounts.

These approaches allow organisations to get ahead of threats, to understand if there is any purposeful activity to foment online hatred that could potentially disrupt users or audiences, and to provide real-time support to individuals exposed to threats, harm or abuse on their platforms.

Supporting users to reduce the risk of harm

A number of companies have programmes focused on empowering users. They report that providing users with the tools to navigate the online environment has benefits such as boosting confidence in participating online, reinforcing positive behaviours or reducing the likelihood of engaging in negative behaviours that exacerbate toxicity. For example, one gaming and entertainment organisation designed a feature that teaches users the most effective way to communicate with online teammates to avoid unintended miscommunications. By empowering individuals to engage in online spaces and raising awareness of trust and safety issues, organisations can help to create safer online behaviours and better digital citizenship.

Several ISPs have funded initiatives to raise awareness of trust and safety issues, including funding of Internet Matters, which provides advice to parents, children, and other users on online safety practices. BT offers interactive games and training to help people protect themselves, their families and their business by helping them learn about online behaviours and detect fraud and scams. BT also formed 'Hope United', a team of top footballers combining to provide advice and tools on how to counter online hate. Similarly, EE established a Digital Champions programme, where staff members visit communities to run sessions about the internet to help support people to get online safely. They also provide information and advice about online security, etiquette online and how to keep children safe.

Protecting brand reputation

There are reputational benefits to positive experiences online, where users feel safe. Stakeholders described how creating safe and trustworthy experiences can lead to competitive advantages, attracting users to a platform and building brand loyalty. This can result in commercial benefits as highlighted in a recent Crisp / Forrester report which emphasised how a return on investment can be achieved by investments in brand safety and threat intelligence.

At the same time, stakeholders highlighted how there are reputational risks from not prioritising trust and safety. In the retail sector, a lack of robust age or identity verification can result in real-world harm and potentially legal consequences. Not only does this impact a company's relationship with the individuals directly involved, but it can also have broader consequences for associations with their brand.

> **"Businesses post-pandemic have a lot more tech, they're more digital and have more of an online presence. Tech adoption … will be more prevalent, building trustworthy platforms and using positive online experiences as a competitive advantage."**
>
> Retail and services economy organisation

In the advertising space, companies are starting to recognise the need to understand advertisement placement, and how this can be controlled in line with brand reputation, values, and responsibilities. This has started to inform principles of 'adjacency standards and control' to set consistent industry standards for placement of advertisements. For example, Global Disinformation Index (GDI) is collaborating with Oracle to provide independent risk rating analysis to support marketers in safeguarding their ad spend and protect their brand from inadvertently supporting sites that host disinformation or harmful content.

Using inventories of domain quality and risk-rating to inform advertisers about problematic ad slots has a number of benefits. First, it encourages organisations to prevent potentially harmful yet unintended consequences to citizens of all ages, communities and organisations. Secondly, it gives organisations the platform intelligence they need to avoid posting adverts that inadvertently fund hate groups or disinformation. And finally, the intelligence can be used to defund harmful platforms and sources of harmful content, protecting organisations' brand reputation and further mitigating the negative impacts to individuals and communities.

"[In 2020], a bunch of advertisers withdrew after seeing their advertising appearing next to hate filled racist content … It won't be sustainable for investors to invest in platforms that don't meet government regulations and societal norms."

Safety tech organisation

"The cost of [engaging] a person in an online digital space is high … Once you get people, you want to keep them along for the journey … Users will only tell you what they need if they feel safe and trusted."

Social media, apps and platforms organisation

Strengthening business operations

Achieving safer online experiences can also deliver direct value to business operations, through enabling greater understanding of user need, setting a corporate ethical direction, and supporting transparency and compliance.

Understanding users

Creating safe and trustworthy online spaces is key to gaining a better understanding of a business' user base, as it opens up the potential for users to share more of their feelings and insights online. This allows businesses to strengthen their products, develop innovations and target their strategies more effectively.

Understanding how to build safer and more trustworthy online experiences does not just have brand benefits, but broader social and democratic ones. When users disengage with the online environment as a result of harmful online experiences, this can also have a knock-on impact on civic participation. For example, research shows that experiencing harms has a range of negative impacts both offline and online. This includes reduced user participation, reduced internet usage, reduced self-esteem or self-confidence and the avoidance of social media.

Organisational values

In addition to building relationships with users, some businesses are prioritising trust and safety as an intrinsic part of their organisational culture and mission.

Companies such as Bumble or Fluttr emphasise the importance of creating safe spaces to develop both online and offline relationships, seeing this as a core part of their service or a unique selling proposition (USP). Similarly, the LEGO Group seeks to put safety at the heart of its offline products and has looked to carry over this ethos to its online products and services.

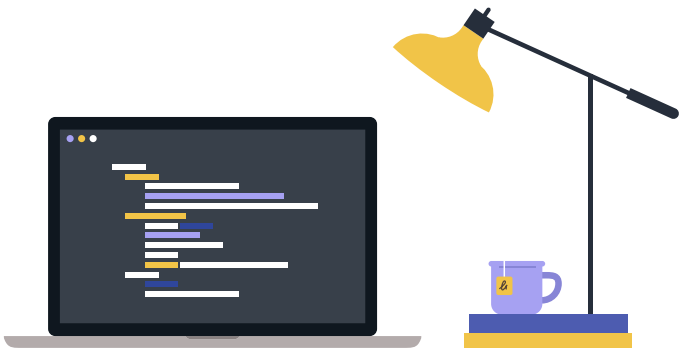
Of course, not all organisations have the same objectives, structures, culture or values, and this has implications for the scalability and transferability of certain approaches. However, the increasing importance of organisational values to both users and businesses demonstrates the potential link between a company's mission and its competitive growth.

> **"We have got a definition of safety, we have got values that spring from the company values, we have got principles … everything that we do related to safety has to align with these things, otherwise we shouldn't be doing it."**
>
> Social media, apps and platforms organisation

Staff retention

Demonstrating a values-driven approach may be critical for staff retention as well as employee wellbeing. We heard from interviewees how demonstrating company values brings corresponding benefits in terms of recruitment savings, productivity and reputation across the digital economy. Stakeholders described how individuals want to work in roles and organisations they believe in and getting the right approach to content moderation or tacking disinformation is becoming increasingly important in a competitive market.

This reflects research which finds that job retention rates in the UK safety tech sector are higher than the tech sector average, suggesting that once people start working in the trust and safety field, they are motivated to stay. Similarly, research carried out by Doteveryone found that almost one in five tech workers in the UK who felt that decisions about a technology were potentially harmful went on to leave their company as a result.



> **"[Values] are something that very specifically potential employees look for when they are choosing a company to work for. So the values a company has and how it lives out those values and what the purpose of a company is, is really important for everybody now when they are looking to change roles in a way that it probably wasn't twenty years ago."**
>
> Technology organisation

Transparency and compliance

Many social media companies use their terms and conditions to set out how they expect their platforms to be used. In turn, these terms and conditions are influenced by the expectations of users, who want to have positive experiences online.

> **"Moderation is about looking at content, leaving or removing it. Community management is about educating members of the community, creating and enforcing community guidelines, and encouraging participation in line with those guidelines."**
>
> Social media, apps and platforms organisation

Some platforms have been relatively open in revealing the complexity of decisions related to content moderation. This openness may help to identify potential problems and find solutions. For example, Twitter responded to user feedback related to potential bias in its image-cropping algorithm by publishing a blog post discussing the issue in detail, and setting out the next steps it would take to address the problems. Twitter went on to change how it displays photos to give greater power to users, rather than relying on machine learning.

Embedding trust and safety in processes can also help businesses to get ahead of the regulatory curve. This has the potential for long-term efficiency and cost savings by avoiding the need to interrupt user experiences or 'retrofit' solutions when new laws are introduced. A number of organisations are starting to approach trust and safety as a positive asset or are building in safety by design principles from the design stage onwards. While it may take time for legislation and regulation to change on a global scale, there are opportunities for the UK economy in establishing benchmarks, exporting solutions, and working with organisations abroad to enhance trust and safety.

"The importance of trust and safety will increase. We expect a big shift in the next five years to child safety by design products … and regulations around the world that enforce child safety by design."

Safety tech organisation

The role of the public sector in supporting the trust and safety ecosystem

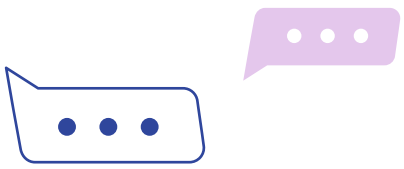
The contribution that online trust and safety can make to positive public policy outcomes is becoming recognised globally. For example, the UK government has played a key role in supporting the wider trust and safety ecosystem through its promotion of the potential of the safety tech sector in national and global fora.

The G7 Internet Safety Principles, published during the UK's 2021 G7 presidency, included a commitment to share information, research and good practice for the development and adoption of safety tech between countries. The UK Online Safety Bill, introduced to Parliament in March 2022, is also expected to be a major driver of trust and safety measures, alongside the UK government's Safety by Design principles. Similar approaches are taking place globally, with the Australia eSafety Commissioner also issuing detailed guidance and advice on how a 'Safety by Design' approach can help embed safety into the culture and leadership of an organisation, and create more positive, civil and rewarding online experiences for everyone.

This focus on trust and safety is also relevant to data protection. In the UK, the Information Commissioner's Office (ICO) is developing standards for online services to protect children's data online, as set out in the Age Appropriate Design Code. It is also funding innovators through their regulatory sandbox programme to explore how services working with high-risk data or novel technologies can comply with data sharing regulations.

"Every business model is different. So we need some flexibility in the solutions, but companies need clear rules and ways to evaluate the solutions … They want a guarantee that it works, but currently a common system of standards to evaluate effectiveness of solutions is lacking."

Retail and service economy organisation



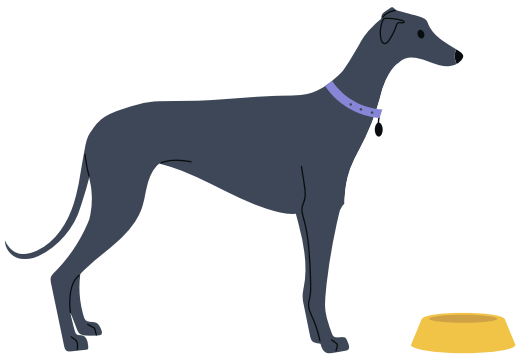
Understanding the value of trust and safety

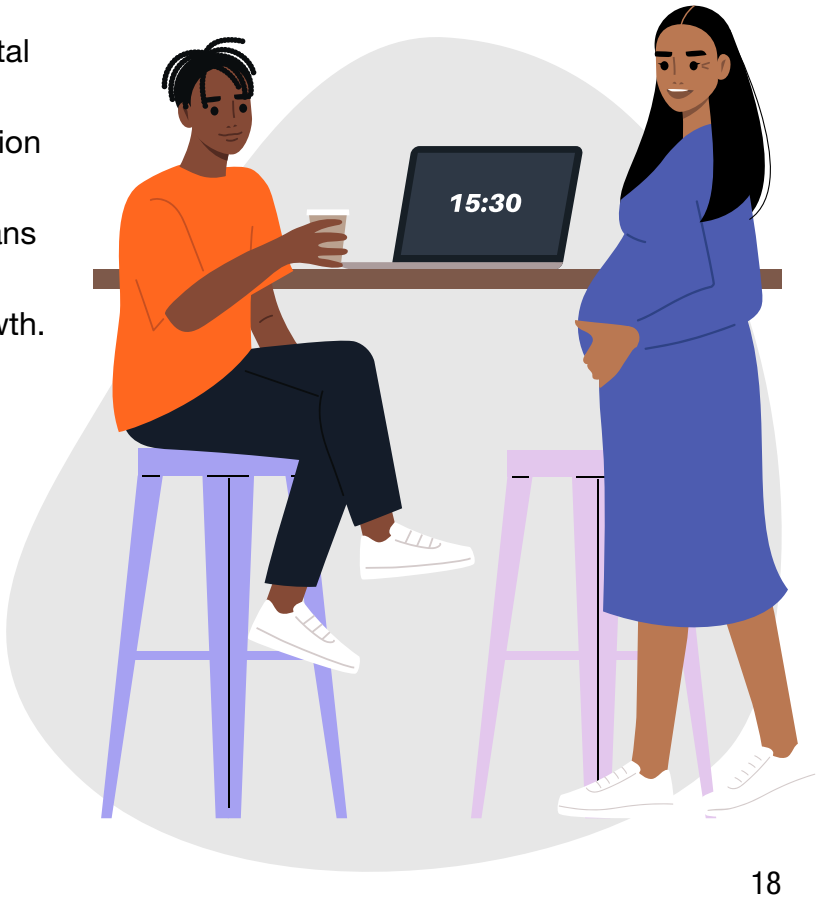
Trust and safety measures could help protect a wide range of sectors in the UK economy. We estimate these sectors collectively contribute £243 billion of value (GVA) to the UK economy.

"With recognition, feasibility, advancement of technology and legislation, the [cyber security] industry exploded. We're at the same point – online safety is just about people, not assets. Online safety could be the next cyber security industry in terms of scale, diversity and importance."

Safety tech organisation

Our research has identified six key sectors within the UK economy which could benefit from the use of safety tech and trust and safety approaches more generally. These include gaming and entertainment; social media, apps and platforms; media and creative industries; retail and services; digital infrastructure; and the public sector. We estimate these sectors contribute £243 billion in GVA each year for the UK economy. The increased digitisation of these sectors means there is a key role for ensuring online trust and safety and promoting sustainable growth.





Trust and Safety: Key Economic Sectors

Sector	Estimated market size	Example use cases for trust and safety
Gaming and Entertainment	At least £24bn in GVA[3]	• Gaming platforms addressing toxicity in online games • Sports clubs using threat and bad actor intelligence to tackle hate, racism and antisemitism in sport • Children's game producers employing age-appropriate design principles
Social Media, Apps and Platforms	There is no formal market estimate for the size of the social media industry in the UK with respect to GVA	• Platforms designing, implementing and enforcing terms and conditions • Technical approaches to content moderation at scale • Encouraging quality discussion, civil engagement and inclusion of marginalised groups in public life
Media and Creative Industries	At least £25bn in GVA[4]	• Advertising and marketing organisations' awareness of 'adjacency control' and inadvertently funding hate groups or disinformation • Media organisations implementing content provenance, authenticity, fact-checking and media literacy campaigns to improve public trust

3. Estimate includes DCMS (2019) Economic GVA Estimates for Sports (£16.9bn), Gambling (£8.3bn), Gaming (£2.9bn), and sector estimates for Children's Toys and Products (£3.3bn – sales proxy) (NPD), and eSports (£111m) (UKIE).

4. Estimate includes DCMS (2019) Economic GVA Estimates for Advertising and Marketing (£17bn) and ONS (2019) measure for news and print media (SIC-based) (£8bn)

Sector	Estimated market size	Example use cases for trust and safety
Retail and Service Economy	At least £159bn in GVA[5]	• Factors such as brand protection, verification of online reviews, verification of goods and services, prevention of sale of harmful goods or services • Impact of technologies (e.g. age assurance) to ensure compliance, reduce cost and improve customer experience • Rise of sharing and gig economies and enhanced demand for user verification in deliveries, transport and hospitality
Digital Infrastructure	At least £35bn in GVA[6]	• Role of digital infrastructure organisations (e.g. telecoms, Internet Service Providers, payment processors) in content takedowns, monitoring, filtering • Disrupting funding of hate groups or disinformation
Public Sector	Our research has identified 69 public contracts awarded for trust and safety related products and services in 2021	• Key buyer of safety technologies

A more detailed breakdown of the economic sector analysis, and related business models, can be found in the appendices to this report.

5. Estimate includes SIC-based measures for GVA for accommodation (£18bn), food and delivery (£42bn) and retail (£100bn)

6. Estimate includes the DCMS Economic Estimates (2019) measure for Telecoms (£35bn)

The costs of online harm

The prevalence and nature of harm can generate real costs for users and businesses. The Online Safety Bill Impact Assessment estimates the annual social cost of online harm (i.e. the combination of direct costs to victims and society, and indirect costs of worsened mental health and lost productivity) to be at least £13 billion per annum. These costs include:

- Child Sexual Exploitation and Abuse (CSEA), which is estimated to cost victims and wider society nearly £5 billion per annum. Of this, the estimated cost of CSEA with an online element is at nearly £1 billion;
- Cyberstalking has an estimated 300,000 victims each year (2020 estimates), which costs the UK economy at least £10 billion per annum;
- The impacts of cyberbullying on victims and the cost of treatment are estimated to cost in excess of £600 million per annum;
- Online intimidation of public figures such as MPs costs almost £5 million per annum through the cost of security measures, police time, and impacts on mental health

Misinformation can also drive real-world costs. For example, it is estimated that misinformation relating to COVID-19 and mask-wearing could have weakened the UK economy by £3.6 billion between Q2 and Q3 2020 through increased caseloads and hospitalisations.

The safety tech market

The DCMS Safety Tech Sectoral Analysis (2022) highlighted a rapidly-growing UK safety tech market, with the sector experiencing an annual 21% revenue growth (to £381m) and 30% employment growth (to 2,850 FTE) across 117 companies. A separate report by Paladin Capital Group on the US safety tech market found evidence of a growing US safety tech sector, whose 160 companies employ 8,800 safety tech professionals. It estimates the sector has already raised over $1bn in external investment (a quarter of this in 2020 alone).

Safety tech, and related services, contains a number of distinct sub-markets, which can be 'pure-play' e.g. content moderation, or more diversified e.g. identity verification, of which age assurance may be one component, or threat intelligence which may cover cyber security techniques. We have identified a high-level overview of some of the current global market estimates (in dollars) that are available:[7]

7. These are high-level global market estimates collated by external agencies such as Gartner, and may contain overlap. These are provided to give an indication of market scale.

- Social media management – $14.4 billion[8]
- Content moderation – $11.8 billion[9]
- Identity verification – $7.6 billion[10]
- Threat intelligence – $5.5 billion (in 2019)[11]
- Digital forensics market – about $4 billion[12]
- Filtering – $4 billion (in 2020)[13]
- Counter-disinformation – (no size provided)
- Kidtech advertising – $1.7 billion[14]
- AI image recognition market – $1.7 billion (in 2020)[15]
- Parental control – $900 million (in 2020)[16]

This highlights the range of technologies available to help protect users online, and the market opportunities for providers to grow and scale. The increase in demand for safety tech products and services is one proxy for understanding how demand will continue to grow as new regulation comes into force.

The professionalisation of trust and safety - developing trust and safety teams

Demand for trust and safety roles in the UK almost doubled in 2021, with more than 10,000 relevant job postings advertised online.

An increasing number of UK businesses are also building multi-disciplinary trust and safety teams in-house, drawing on the expertise of a growing number of trust and safety professionals. This is a further sign of a maturing sector.

Building in-house trust and safety teams, often coupled with the deployment of safety tech solutions, allows businesses to both develop bespoke technical solutions and consider the wider norms and values they want to embed in their approach to trust and safety. Many teams are increasingly multi-disciplinary and being

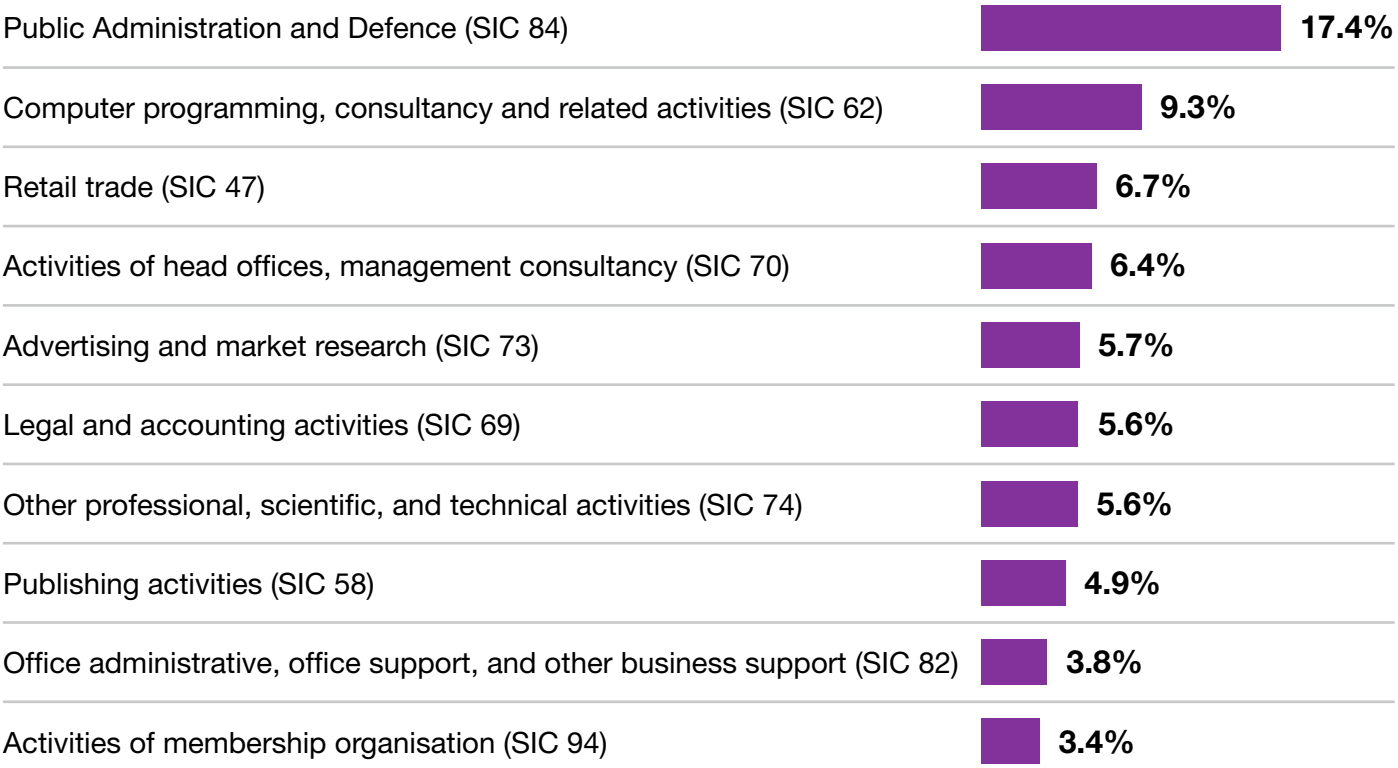
8. ResearchAndMarkets.com, Global Social Media Management Market (2021 to 2026) - Enhancement of Customer Experience With Social Media Management is Driving Growth, 2021
9. Transparency Market Research, Content Moderation Solutions Market Study, 2020
10. marketsandmarkets.com, Identity Verification Market
11. verifiedmarketresearch.com, Threat Intelligence Market Size And Forecast
12. marketsandmarkets.com, Digital Forensics Market
13. verifiedmarketresearch.com, Web Filtering Market Size And Forecast
14. PwC and SuperAwesome, Kids Digital Media Report 2019
16. Mordor Intelligence, AI Image Recognition Market
17. Fortune Business Insights, Parental Control Software Market Size, Share & COVID-19 Impact Analysis, 2022

led by experienced professionals, who bring an understanding of how social platforms work and can be improved from a trust and safety perspective.

Our analysis of online job postings in the UK shows that the **demand for trust and safety professionals almost doubled in 2021**, with 10,293 job postings linked to trust and safety in the UK, compared with 5,284 job vacancies identified in 2020. Roles span social media strategists and moderators, software developers

> **"Now we have experienced professionals. We have the luxury to build on a class of professionals who have learnt from our mistakes. As they come with seniority, they will change things. The people who have had that journey, they now run the show and we are doing it very differently from the beginning."**
>
> Safety tech organisation

Figure 1.1: Top 10 industry sectors hiring for trust and safety affiliated roles (based on two-digit SIC classifications)



Source: Perspective Economics analysis of online vacancy data (n = 4,019 with a known employer and sector)

and engineers, (online) community managers, data analysts, security engineers, safety and product managers, and legal or policy specialists.

Demand is also not constrained to a single sector, but spans public roles, retail, consultancy and advertising as well as technology related activities.

For each of these sectors, some notable employer types for trust and safety roles include:

- **Safety tech** companies recruiting for content moderation specialists, UX designers, and software engineers, data scientists, threat intelligence analysts, and language skills
- **Large social media platforms** building internal capacity with respect to online harms research, legal counsel, data science, moderation and safety functions, and brand safety
- **Games developers** recruiting for brand management, community leads, and technical leads for trust and safety
- **Media organisations** hiring for content management and identification, and response to disinformation
- **Large retailers** building social media, brand safety, and community moderator functions, with safety by design principles in place
- **Police forces** recruiting for digital forensics investigation and technical roles, anthropologists and psychologists
- **Regulators and public bodies** such as Ofcom, DCMS, and the Home Office are all building teams to understand how to identify, regulate, and manage online content, and how to design public policy to address online harms

Online job boards such as LinkedIn and Indeed also provide insight into the demand for trust and safety professionals. For example, an initial review of LinkedIn suggests there are c. 17,000 profiles referring to 'trust and safety' internationally, spanning employers such as large social media platforms and retailers. In addition, in the United States, initiatives such as the Trust and Safety Professional Association (TSPA) have established their own job board with employers such as TikTok, Roblox, Airbnb, and Pinterest.

Conclusion

Growing online trust and safety is not an easy job, but – as this report shows – good work is already happening. The UK safety tech market is one of the fastest growing in the UK digital economy and there were more than 10,000 job postings related to trust and safety in the UK in 2021, almost double the previous year.

We have identified six key sectors that could benefit from safety tech and trust and safety approaches. These are gaming and entertainment; social media, apps and platforms; media and creative industries; retail and services; digital infrastructure; and the public sector. Collectively, we estimate that these sectors contribute £243 billion of value (GVA) to the UK economy.

But a better digital world will create more than economic advantages – it will be better, safer and more trustworthy for everyone. The pioneers of trust and safety are out there, delivering value to their companies while ensuring their online audiences are part of healthy, nurturing and supportive communities. This will create dividends for business and the whole of society by creating better, safer and more trustworthy online experiences.

The benefits include:

- reducing the risks and costs of online harm
- greater ethical and legal compliance
- retaining staff and users
- protecting and enhancing businesses' brand and reputation
- driving online and offline user engagement
- generating better insights about users
- aligning organisational values with user expectations

Our research has shown that building a better and safer internet is possible, driven by a stronger, more sustainable digital economy in which companies do more to recognise the value of people's safety and wellbeing.

Our standards and accreditations

Ipsos' standards and accreditations provide our clients with the peace of mind that they can always depend on us to deliver reliable, sustainable findings. Our focus on quality and continuous improvement means we have embedded a "right first time" approach throughout our organisation.

ISO 20252

This is the international market research specific standard that supersedes BS 7911/MRQSA and incorporates IQCS (Interviewer Quality Control Scheme). It covers the five stages of a Market Research project. Ipsos was the first company in the world to gain this accreditation.

Market Research Society (MRS) Company Partnership

By being an MRS Company Partner, Ipsos endorses and supports the core MRS brand values of professionalism, research excellence and business effectiveness, and commits to comply with the MRS Code of Conduct throughout the organisation. We were the first company to sign up to the requirements and self-regulation of the MRS Code. More than 350 companies have followed our lead.

ISO 9001

This is the international general company standard with a focus on continual improvement through quality management systems. In 1994, we became one of the early adopters of the ISO 9001 business standard.

ISO 27001

This is the international standard for information security, designed to ensure the selection of adequate and proportionate security controls. Ipsos was the first research company in the UK to be awarded this in August 2008.

The UK General Data Protection Regulation (GDPR) and the UK Data Protection Act (DPA) 2018

Ipsos is required to comply with the UK GDPR and the UK DPA. It covers the processing of personal data and the protection of privacy.

HMG Cyber Essentials

This is a government-backed scheme and a key deliverable of the UK's National Cyber Security Programme. Ipsos was assessment-validated for Cyber Essentials certification in 2016. Cyber Essentials defines a set of controls which, when properly implemented, provide organisations with basic protection from the most prevalent forms of threat coming from the internet.

Fair Data

Ipsos is signed up as a "Fair Data" company, agreeing to adhere to 10 core principles. The principles support and complement other standards such as ISOs, and the requirements of Data Protection legislation.

Perspective Economics

Careful Industries

Ipsos