GOV•UK

Department for
Science, Innovation
& Technology

Research and analysis

Deepfake detection technology

Published 26 March 2026

Contents

This publication is available at https://www.gov.uk/government/publications/deepfake-detection-technology/deepfake-detection-technology

Executive summary

What is in this report?

Deepfake detection technology has emerged as a critical field in recent years, combating the growing threats posed by deepfakes, a subset of synthetic media involving audio-visual content that is generated or manipulated using AI to misrepresent someone or something.

Despite progress being made through initiatives such as the UK government's Deepfake Detection Challenge and ongoing research efforts in AI identification, particularly within the policing and law enforcement sectors, the development of deepfake detection technology remains in its early stages, and the market itself remains nascent.

In late 2024, the Department for Science, Innovation and Technology (DSIT) commissioned PUBLIC to investigate the state of the UK deepfake detection market and its potential for growth. Through a combination of a literature review, expert engagement, and provider mapping, this report aims to establish a robust evidence base to inform how the government can support the continued growth of this market.

This report covers:

- the current state of demand and supply in the UK deepfake detection market
- existing barriers and key drivers shaping the future of the sector
- an analysis of the future evolution of the deepfake detection market, with a focus on achievable future scenarios, and recommendations for the necessary changes to achieve those outcomes

This report was prepared by a specialist digital advisory firm, PUBLIC Group International Ltd, who have been supporting DSIT with this analysis. This report is not official government policy.

Key findings

The mitigation of deepfake threats occurs at multiple stages: before, during and after their creation. This report specifically focuses on deepfake detection, which encompasses a range of technologies or solutions designed to reveal the origins of content, regardless of whether information

has been attached to it. The deepfake detection market is still in its nascent stages, with increasing adoption across sectors despite significant barriers to market entry, adoption and scaling.

Demand. Deepfake detection is becoming critical across sectors for various reasons. Businesses are increasingly integrating deepfake detection for fraud prevention, brand protection, identity verification and content moderation. Governments and public authorities are leveraging these technologies to enhance national security, combat misinformation and disinformation, and support law enforcement investigations and prosecutions.

Supply. Since 2017, the global deepfake detection market has experienced rapid growth, led by 23 US-headquartered providers out of 59 identified third-party firms as of 2025, followed by 7 from the UK, with the remainder spread across other regions. Many of these providers remain in the pre-seed or seed stages of funding, with an average total funding of £25 million. The primary technical approaches focus on machine learning, particularly neural network architectures and feature-based detection methods.

Barriers. Key barriers to entry and adoption include:

- An evolving international online safety regulatory landscape creating uncertainty for both suppliers and customers
- High technical costs and resource constraints, leading to a low perceived return on investment
- Concerns over detection reliability, causing hesitancy among potential customers and the general public
- Limited representative training data, making detection systems vulnerable to novel, real-world manipulation
- Variability in accuracy metrics and testing datasets, which hinders meaningful performance comparisons.

Drivers. The future market direction will depend on the degree to which the following factors are present:

- Rapid advancement and widespread availability of GenAI, increasing the demand for advanced detection solutions
- Concerns about the role of AI in national security and public safety from crime, driving political and policy attention
- Development of clear regulatory frameworks and enforcement mechanisms, providing certainty for the market
- Improving customer understanding and changing user behaviours in response to deepfake threats

- Favourable market entry conditions for foreign vendors, positioning the UK as an attractive market for innovation
- Better access to high-quality training data, enabling the development of more robust detection solutions
- Standardised accuracy testing that is essential for technology maturity and building buyer confidence.

1. The Context for Deepfake Detection

> In 2025, the Department for Science, Innovation and Technology (DSIT) commissioned PUBLIC to investigate the state of the UK deepfake detection market and its potential.

> Through a literature review of over 80 sources, 14 expert interviews, a workshop, and the mapping of 59 deepfake detection providers in the UK and globally, this report aims to develop a robust evidence base and inform how the government can support the continued growth of this market.

In late 2024, DSIT commissioned PUBLIC to investigate the state of the UK deepfake detection market and its potential

From December 2024 to March 2025, the Department for Science, Innovation and Technology (DSIT) commissioned PUBLIC, a specialist digital advisory firm, to undertake deep-dive research into the current state of the UK deepfake detection market to inform how the government might incentivise and grow this market.

Challenge
Deepfake detection technology is rapidly evolving to combat threats posed by AI-generated content. While progress is being made through government initiatives, industry collaboration, technological solutions, and ongoing research, the effectiveness of detection tools remains a challenge due to the rapid evolution of deepfake technology.

Project Objectives

1. Develop a **robust evidence** base on the UK deepfake detection market, analysing the **market's maturity, trends** and the **UK's global positioning** in deepfake detection.
2. Identify **barriers** to adoption and scaling of deepfake detection solutions and **key factors driving future growth.**
3. Explore the **future evolution** of the market and **HMG/DSIT's roles** to support continued growth of the UK deepfake detection market.

Deepfakes are AI-generated media that misrepresent reality and can have the potential to cause harm

While there is no universally accepted definition of deepfake, this report adapts Ofcom's definition to include unintended deepfakes.

Synthetic Media
Video, image, text or audio that has been generated in whole or partly by Artificial Intelligence (AI) algorithms.

> **Deepfake** is a subset of synthetic media.
>
> Audio-visual content that has been generated or manipulated using AI, and that misrepresents someone or something.

Characteristics of deepfakes:

1. **Created using AI**, rather than simpler methods
2. **Audio-visual content**, rather than text
3. **Misrepresenting someone or something** Deepfakes may depict real or fictitious people, events, or objects
4. **Potential to cause harm.** Deepfakes can have the potential to cause harm, regardless of intent

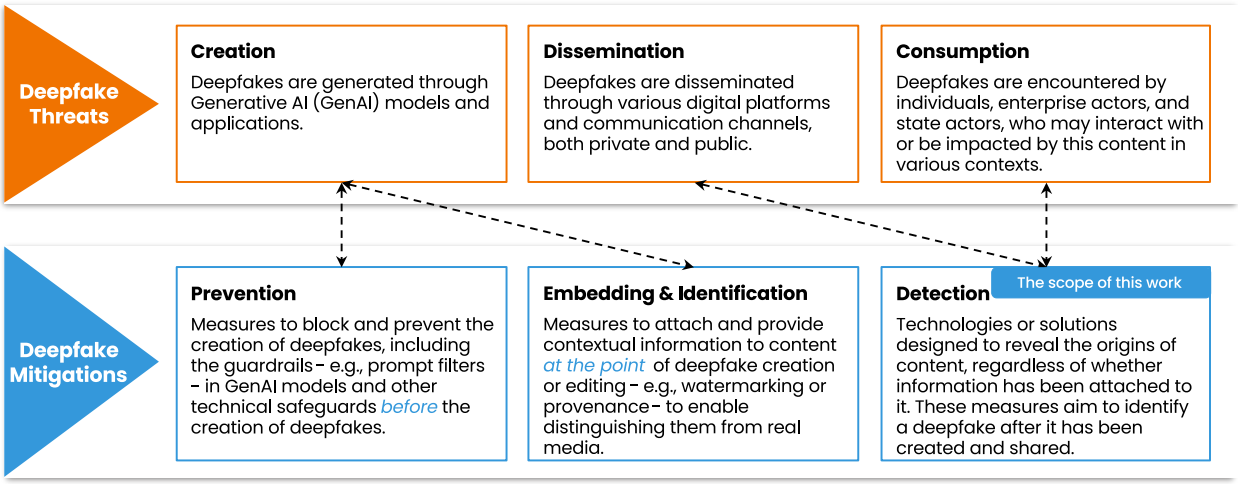
Sources:

(1): DRCF, The Future of Synthetic Media (https://www.drcf.org.uk/siteassets/drcf/pdf-files/the-future-of-synthetic-media.pdf?v=385978), 2024.

(2): Ofcom, Deepfake Defences: Mitigating the Harms of Deceptive Deepfakes (https://www.ofcom.org.uk/siteassets/resources/documents/consultations/discussion-papers/deepfake-defences/deepfake-defences.pdf?v=370754), 2024. We adapted the definitions of synthetic media and deepfakes based on Ofcom's definition.

Emerging technologies and solutions countering the creation, dissemination and consumption of deepfakes

In particular, this report focuses on **deepfake detection.**



This report aims to answer five research questions

1 **What is the current state of maturity of the deepfake detection market in the UK, and how has it evolved with AI capability development? How is it expected to evolve in the face of international competition?**

Analysing the supply and demand in the UK deepfake detection market, including its composition, segmentation, growth, primary use cases in both public and private sectors, to assess the maturity of detection technology and the market.

2 **What accuracy metrics are currently used to evaluate deepfake detection tools? What is the most sensible approach to measuring accuracy?**

Collating various accuracy metrics and key considerations for evaluation, to support the understanding of the efficacy of detection tools and the potential role of the government moving forward.

3 **What are the key barriers and drivers to the commercial success for third-party deepfake detection providers?**

Performing a PEST analysis to legislative and regulatory, commercial, social and technological factors that are either slowing down or driving market entry and deployment.

4 **How does deepfake detection technology interface with other aspects of the safety tech industry?**

Contextualising the deepfake detection market within the broader safety tech ecosystem to highlight the evolution of the trust and safety landscape.

5 **How can the Government support continued growth of the UK deepfake detection market?**

Identifying the role the government can play in addressing existing barriers between deepfake detection demand and supply, and in preparing for the future.

The findings and recommendations are based on robust evidence and a structured research methodology

Evidence Gathering

We conducted a comprehensive review of diverse sources on the deepfake detection market landscape.

- **86** - Literature reviewed and referenced in the report
- **14** - Expert & stakeholder interviews conducted
- **59** - Deepfake detection providers mapped

Trends Analysis

From the evidence base, we identified key market trends, barriers and drivers, using the PEST framework.

- **7** - Primary use cases mapped
- **9** - Drivers identified that are shaping the future

Scenario Development

We then identified the best possible future for the market, and the pathway from its current state to that vision.

- **1** - Future workshop conducted
- **2** - Scenarios developed (1 current state, and 1 future state)

Recommendations

Finally, we developed recommendations on where government action or further work by DSIT would be impactful.

- **5** - Recommendations on the roles of the government

2. The Current state of UK Deepfake Detection Market

> The deepfake detection market is still in its early stages, with increasing adoption in key use cases such as fraud prevention, identity verification, content moderation, national security, and law enforcement.

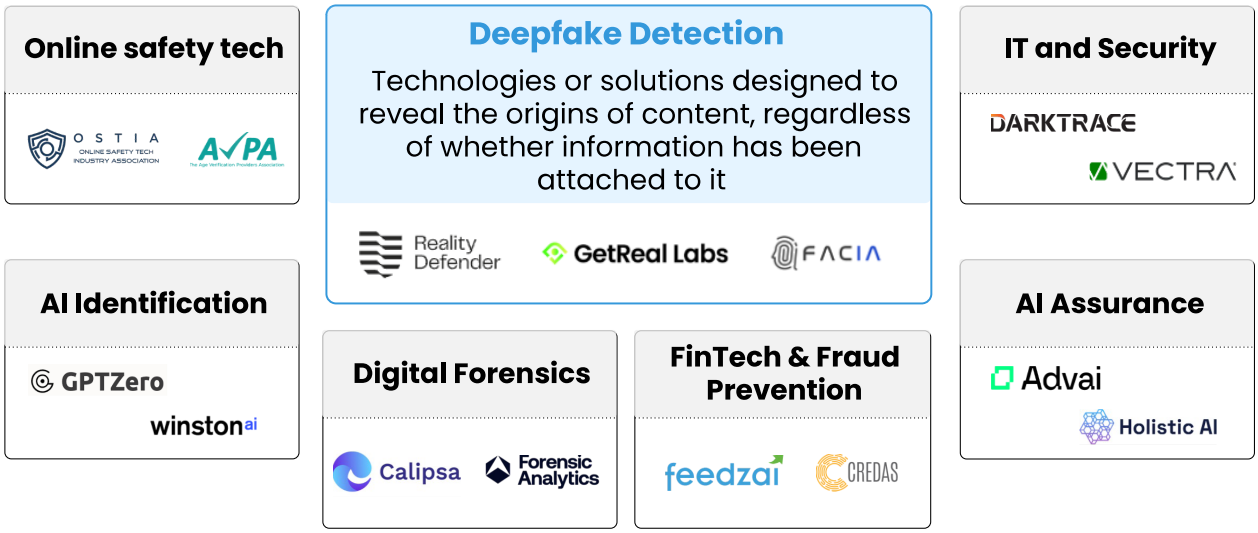
> Despite notable growth since 2017, led by US and UK players, most providers remain in early-stage funding rounds and struggle to scale.

> Key barriers to entry and adoption include regulatory uncertainty, high costs of technological development, resource constraints, low perceived return on investment (ROI), limited confidence in detection accuracy, and lack of training data.

Overview

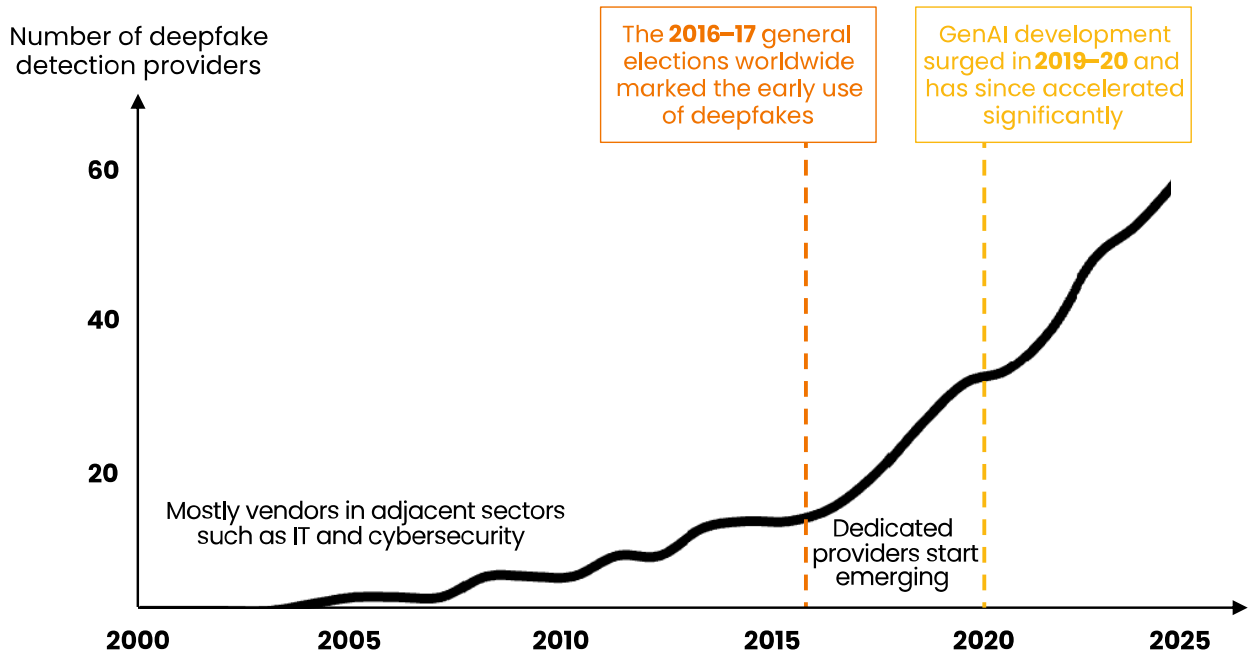
Deepfake detection is a crucial aspect of the online trust and safety ecosystem and AI governance

Detection and its adjacent sectors



The global deepfake detection market is in the early stages, with a surge in market growth since 2020

The deepfake detection market is expected to **continue growing and diversifying** in terms of use cases over the next few years, driven by several factors outlined in the next section, although **the pace may slow.**



Notes: (1): The indicative market growth trend (in terms of the number of providers founded) is based on the analysis of 59 providers identified in this research.

Demand Analysis

Deepfake detection is being adopted across various sectors to combat growing threats

Seven primary use cases were identified as part of this research

Fraud Prevention & Cybersecurity	Misinformation, Disinformation & Narrative Manipulation Detection	Identity & Age Verification	Reputation, Branc Protection & Social Monitoring
Detecting deepfake-enabled identity fraud, financial scams, cyber threats, and account takeovers	Identifying manipulated media used for political deception, misinformation, and social manipulation	Preventing synthetic identity fraud, verifying age and identity for digital access, and securing biometric authentication systems	Detecting and mitigating deepfakes that misrepresent individuals, brands or organisations, manipulate public perception, or threaten reputation

Fraud Prevention & Cybersecurity	Misinformation, Disinformation & Narrative Manipulation Detection	Identity & Age Verification	Reputation, Brand Protection & Social Monitoring
Fraud & Cybercrime	**Mis/Disinformation**	**Cross-cutting**	**Mis/Disinformation & Fraud**

However, the adoption is constrained by technical, legal and ethical concerns

Below are the key considerations for each of the seven primary use cases

Fraud Prevention & Cybersecurity	Misinformation, Disinformation & Narrative Manipulation Detection	Identity & Age Verification	Reputation, Brand Protection & Social Monitoring	Content Moderati
Evolving fraud tactics that require constant updates to detection tools	Real-time detection and response to manipulated media	Privacy and ethical concerns in handling and using sensitive biometric data	High risk of deepfakes going viral before they can be detected and addressed	Handling the volum of conten to moder effectivel
Scalability challenges in processing high volumes, real-time fraud detection	Countering coordinated narrative manipulation across platforms		Reputational risk demands contextual, not just automated, analysis	Ensuring real-time detection harmful content

Fraud Prevention & Cybersecurity	Misinformation, Disinformation & Narrative Manipulation Detection	Identity & Age Verification	Reputation, Brand Protection & Social Monitoring	Content Moderati
Privacy concerns in implementing deepfake detection tools in biometric systems				Navigatin legal challenge and complyin with globa regulation for conte moderati

The public sector is adopting deepfake detection to protect security, democracy, and public trust

Government, law enforcement, defence, and healthcare agencies are increasingly investing in detection tools to verify evidence, combat fraud, and misinformation and threats that harm public safety and democratic integrity

Government

Public sector bodies, regulatory agencies, and electoral integrity bodies require deepfake detection to combat **fraud, electoral manipulation, voice impersonation, and the interference of public service delivery.** These tools are critical for safeguarding democratic processes, maintaining public trust, and ensuring the integrity of civic functions.

Defence & National Security

This sector requires deepfake detection to counter mis/**disinformation, system infiltration, intelligence deception, and terrorist threats**, to maintain national stability. National security agencies, military bodies, and intelligence services rely on these tools to identify synthetic media that manipulates public perception, compromises critical infrastructure, and disrupts operations.

Law Enforcement

Law enforcement agencies require deepfake detection to combat **deepfake-enabled crimes, including fraud, blackmail, and child exploitation, as well as to identify fraudulent activities and prevent evidence manipulation.** These tools are essential for maintaining the integrity of

investigations, ensuring fair legal processes, and protecting public safety from deepfake-enabled deception.

Healthcare
Healthcare providers and public health agencies need deepfake detection to **prevent medical fraud, data breaches, and patient deception.** Deepfakes can manipulate diagnostics, impersonate doctors, and spread false medical advice, endangering patient safety. Detection tools safeguard telemedicine, protect sensitive data, and maintain trust in medical communications.

Businesses are integrating deepfake detection to fight fraud and safeguard operations, consumer trust, and security

From finance to media, private sector companies are adopting detection tools to prevent scams, financial losses, and reputational damage while countering synthetic media threats that erode security and consumer confidence

Banking, Financial Services & Insurance (BFSI)
BFSI institutions require deepfake detection to combat **sophisticated fraud, identity theft, voice cloning, and synthetic identity fraud.** These tools are essential for securing authentication, detecting manipulated claims, preventing financial scams, safeguarding digital transactions, and ensuring trust in financial services and risk management processes.

Corporate Security & Risk Management
Enterprises and risk management teams need deepfake detection to prevent **fraud, insider threats, and corporate espionage**, while protecting against **reputational damage**. These tools verify identities, safeguard sensitive communications, and maintain operational integrity in the face of manipulated media that could enable scams, data breaches, or reputational damage.

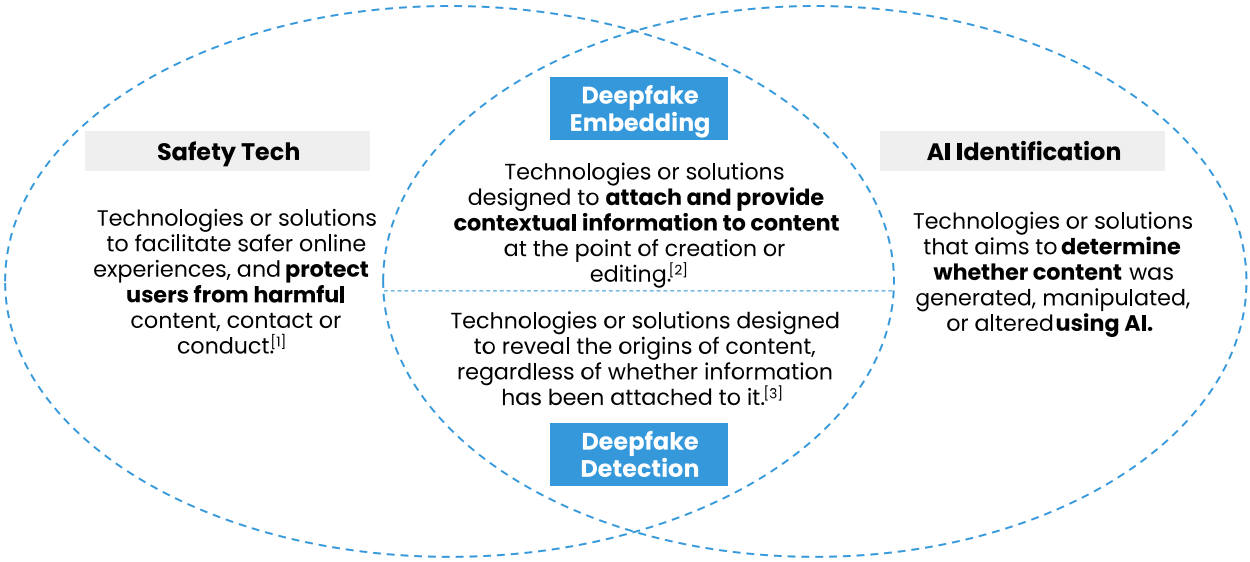
Media & Entertainment
News, media agencies, and entertainment companies need deepfake detection to combat **manipulated content, protect their reputation and brand, and counter misinformation and IP misuse.** These tools are crucial for verifying content authenticity, preventing digital deception, safeguarding creative assets, and maintaining audience trust.

Social Platforms and Content Moderation
Social media platforms and online communities need deepfake detection to combat **harmful content, misinformation, and harassment.** These tools

enforce policies, detect manipulated media, and protect users from deception and digital harm. Automated detection is crucial to maintaining trust and platform integrity at scale.

Supply Analysis

Deepfake detection providers offer solutions in conjunction with online safety tech and AI identification



Note The initial scope of this project focuses on deepfake detection. Since deepfake detection tools are often paired with embedding tools and many providers develop both solutions, both categories are included in the following supply-side analysis.

Sources:

(1): DSIT, The UK Safety Tech Sector: 2024 Analysis (https://assets.publishing.service.gov.uk/media/6707a9f030536cb927482f69/uk_safety_tech_sector_2024_analysis.pdf), 2024.

(2), (3): Ofcom, Deepfake Defences: Mitigating the Harms of Deceptive Deepfakes (https://www.ofcom.org.uk/siteassets/resources/documents/consultations/discussion-papers/deepfake-defences/deepfake-defences.pdf?v=370754), 2024. Definitions adapted from those provided.

Within the realm of deepfakes, most providers focus on detection than on preventative measures

A snapshot of providers

Safety Tech

Tech or solutions to facilitate safer online experiences, and protect users from harmful content, contact or conduct.

AI Identification

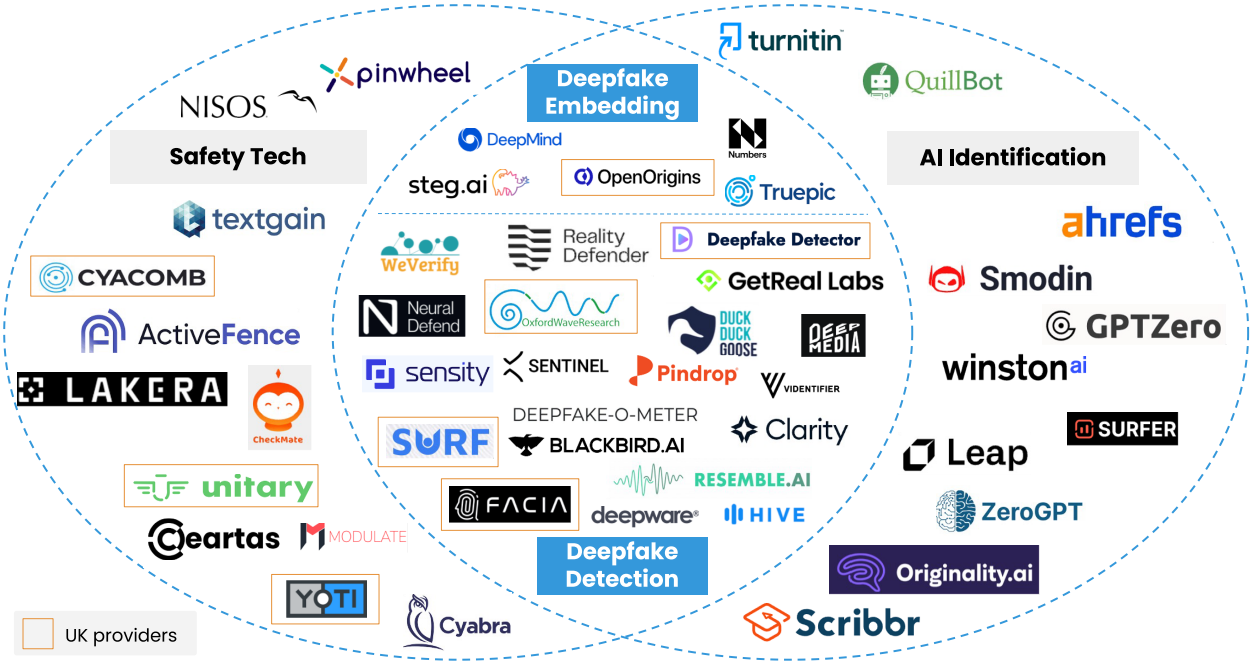
Tech or solutions that aim to determine whether content was generated, manipulated, or altered using AI.

Deepfake Embedding

Tech or solutions designed to attach and provide contextual information to content at the point of creation or editing

Deepfake Detection

Tech or solutions designed to reveal the origins of content, regardless of whether information has been attached to it.



This study mapped 59 deepfake detection providers to capture a snapshot of the global landscape

- **Overall:** The global deepfake detection market is **experiencing rapid growth with the number of providers increasing nearly 380% since 2017** and US-headquartered firms leading the sector. Of the 59 third-party deepfake detection providers identified worldwide, **7 are based in the UK**, positioning it as the second-largest economy in this market.
- **Growth:** While the market began emerging in the early 2000s, dedicated deepfake detection providers only began to surface in 2017. Since then, the market has **expanded by nearly 380%**, largely driven by companies specialising in **machine learning** techniques for deepfake detection.
- **Size:** The majority of providers (83.0%) are **micro or small enterprises.** Dedicated providers tend to be even smaller in terms of employee size and funding compared to diversified counterparts. This is mainly due to their relative youth and their progression through earlier stages of development.
- **Investment:** Many of these providers remain in the **early stages of funding**, with an average total funding of £25 million. Specifically, 8 (36.4%) dedicated providers are in the pre-seed or seed stages, while only one (Reality Defender) is in Series A.
- **Segments:** The providers mapped predominantly focus on **fraud prevention** (72.9%) and **identity and age verification** (45.8%) services to the financial industry, and **mis/disinformation detection** (50.8%) for government, media, or social media platforms.

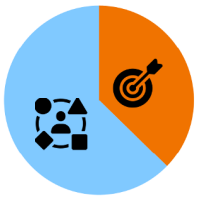
Source: Data on this slide are based on PUBLIC's analysis of 59 third-party deepfake detection providers using open-source research. See Appendix A2 for methodology details.

Global Deepfake Detection Market Snapshot

Detection Techniques

Techniques	Providers
Machine-learning detection	58
Metadata analysis	12
Watermark	2
Hash matching	3

59 Providers in total



37 diversified

22 dedicated

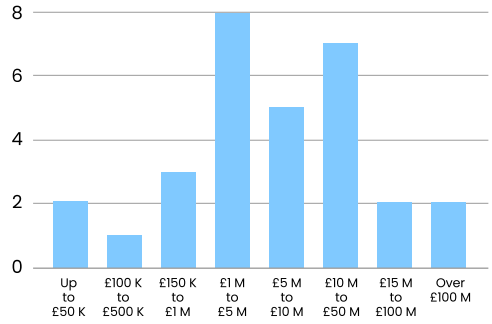
Investment

£25.0m *Avg. total funding per provider*

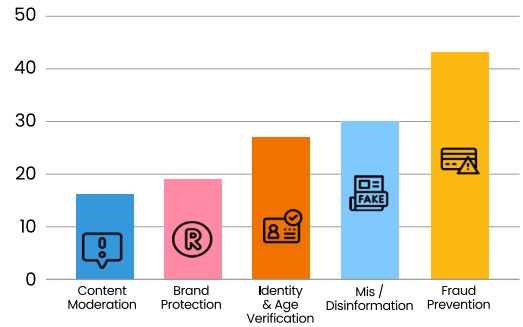
£33.0m *Avg. total funding per diversified provider*

£6.6m *Avg. total funding per dedicated provider*

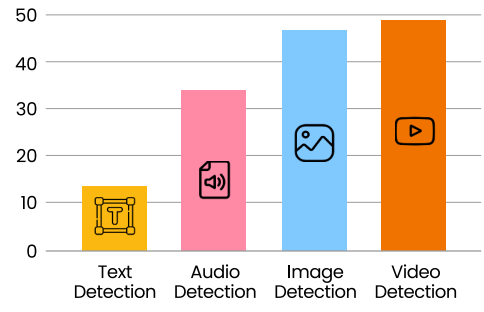
Diversified providers have secured significantly higher funding than dedicated providers.



Use Cases



Media Type



Providers are exploring various way to enter the market and develop their deepfake detection capabilities

Specialist
Sole Focus on Deepfake Detection

These providers dedicate themselves exclusively to the development of deepfake detection technologies, services and solutions, **establishing themselves as specialists** in this field.

Examples:
Reality Defender
GetReal Labs

Expansion
Building on Adjacent Capabilities

Some providers, particularly those already established in adjacent fields, are expanding their service offerings to include deepfake detection. Through **internal R&D and strategic acquisitions**, these providers naturally extend their capabilities to meet the growing demand from clients seeking deepfake detection solutions.

Examples:
SURF
VIDENTIFIER

Partnership
Leveraging External Capabilities

Rather than develop deepfake detection technology in-house, some providers **partner with specialists** in this field, given internal resource constraints. This allows them to integrate advanced deepfake detection tools into their existing services, broadening their offerings without the need for significant technology development investment.

Examples:
Proof (Partnership with Reality Defender)

Detection solutions are expanding partnerships with online safety tech providers and AI developers

Partnership with online safety technology

Deepfake detection providers partner with other deepfake detectors or online safety tech companies to offer comprehensive solutions for detecting and preventing AI-generated threats, combining specialised expertise in deepfake detection with content moderation, identity verification and fraud prevention capabilities.

For example: **DeepTrust** and **Breacher.ai** announced a partnership to combat deepfake AI attacks, offering a joint solution, DeepBreach, for forensic analysis of potentially fraudulent content.(1) **iProov's** liveness detection technology is integrated into **Jumio's** identity verification and KYC platform to combat sophisticated cyber attack using deepfakes.(2)

Partnership with AI developers & GenAI services

Deepfake detection providers partner with AI developer and GenAI services to improve their detection algorithms and accuracy, but also to stay ahead of evolving threats.

For example: **Reality Defender** partners with **ElevenLabs**, an AI voice generator, to advance audio deepfake detection models, and leverage ElevenLabs' proprietary technology to enhance its detection tools.(3) **DeepTrust** partners with **Musicify**, a voice clone and AI music generation service, to join forces on responsible AI usage and development, focusing on AI music detection and IP protection.(4)

Sources:

(1): DeepTrust, DeepTrust and Breacher AI Partnership Announcement (https://www.deeptrust.ai/blog/deeptrust-and-breacher-ai-partnership-announcement-2/), 2024.

(2): Jumio, Jumio Adds iProov's Award-Winning Liveness Detection to its KYX Platform (https://www.jumio.com/about/press-releases/iproov-liveness-detection-kyx/), 2021.

(3): Reality Defender, Strengthening Digital Trust: How ElevenLabs Helps Reality Defender Advance Voice Deepfake Detection (https://www.realitydefender.com/blog/elevenlabs-reality-defender-voice-deepfake-detection), 2025.

(4): DeepTrust, DeepTrust and Musicfy Partnership Announcement (https://www.deeptrust.ai/blog/deeptrust-and-musicfy-partnership-announcement/), 2024.

Big tech companies are developing in-house detection solutions as well as watermarking solutions for identification

Tech giants have developed their own deepfake detection (and identification) solutions, with examples such as:

> Detection
> **Microsoft Video Authenticator(1)**

Microsoft Video Authenticator can analyse a still **image or video** in real time to provide a confidence score that the media is artificially manipulated. It uses **machine learning algorithms** to detect subtleties like pixel inconsistencies and colour fading.

The tool, launched in 2020, is developed In partnership with Reality Defender to combat AI-generated disinformation.

> Detection
> **Intel's FakeCatcher(2)**

Intel's FakeCatcher is a **real-time deepfake detector** that analyses biological signals to determine the authenticity of a video. FakeCatcher uses Photoplethysmography (PPG) - a technique that analyses 'blood flow' in video pixels to determine a video's authenticity in milliseconds.

The tool was first launched in 2022.

> Identification
> **Google Deepmind's SynthID(3)**

SynthID **watermarks** and **identifies** AI-generated content across various media types, including **images, audio, text and video**, by embedding digital watermarks directly into the content **during generation** and using **deep learning** models for subsequent identification.

The Beta version (current version) was launched in 2023 and was made available in open source.

> Identification
> **Meta Video Seal(4)**

Meta Video Seal is a neural **watermarking tool** that embeds durable, invisible watermarks for AI-generated **videos - even after video editing.** The tool also allows for embedding up to a 6-character hidden message within the watermarks.

Video Seal was launched in December 2024, following the release of Audio Seal in June. Video Seal is available in open source.

Sources:

(1): Microsoft, New Steps to Combat Disinformation (https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/?utm_source=chatgpt.com), 2020.

(2): Intel, Real-time FakeCatcher for Deepfake Detection (https://www.intel.com/content/www/us/en/research/trusted-media-deepfake-detection.html).

(3): Google Deepmind, SynthID (https://deepmind.google/technologies/synthid/).

(4): Meta, Video Seal - Meta FAIR AI Demos (https://aidemos.meta.com/videoseal)

Barrier Analysis

Despite the growing demand and supply, deepfake detection adoption has been slow

Barriers to technology development and adoption

Political [Regulatory]
Regulatory uncertainties, delays, and compliance barriers challenge third-party providers despite global legislative efforts.

Regulatory lag and ambiguity

UK market entry and compliance hurdle

Economic [Commercial]
High technical costs, resource constraints, and low ROI perception of T&S remain big barriers to deepfake detection innovation.

Customers perceiving low ROI in deepfake detection

High costs and resource constraints

Social
There is rising public awareness of deepfake threats, but scepticism toward detection tools persists.

End user scepticism toward detection tools

Technological
GenAI and ML advancements drive detection solutions, but limited datasets, inconsistent metrics, and evolving threats remain challenges.

Low access to high-quality training datasets

Heterogeneity of testing datasets and metrics

Despite political focus on deepfakes, regulatory ambiguities and delays remain major challenges

Political / Regulatory trends and barriers

Political deepfake threats eroding institutional trust: Deepfakes diminish institutional trust by enabling the spread of false and damaging mis/disinformation, a concern heightened by 2024's global election cycle. This growing threat drives political and regulatory action to mitigate its risks.[1][10]

Increasing global efforts on legislation against deepfakes: Governments worldwide are ramping up efforts to regulate deepfakes. In the UK, the upcoming Crime and Policing Bill would ban AI-models optimised to produce CSAM. The EU is enforcing transparency requirements through the AI Act and Digital Services Act. In the US, states like California are enacting targeted laws, alongside federal efforts like the NO FAKES Act and Take It Down Act. Spain plans to fine unlabelled AI-generated content and criminalise AI-generated sexual images. Countries like Australia and South Korea have also implemented similar measures.[2][3][4][5][6][7][8][12] These are driving increased compliance demands and regulatory scrutiny.

Barriers

- **Regulatory ambiguity is leaving deepfakes a low priority for online platforms:** Despite growing regulatory momentum, ambiguity remains. While the Online Safety Act (OSA) addresses illegal harm, including both deepfakes and non-deepfakes, the safety tech industry perceives a lack of clarity around its application to deepfakes or synthetic media. This perceived ambiguity prompts platforms to take a reactive, 'wait and see' approach to addressing deepfakes.[7][12]
- **Regulations lagging behind deepfake threats and technology:** AI advancements, including deepfake technology, normally evolve faster than regulatory frameworks can adapt. As a result, regulatory responses currently remain limited, creating uncertainty and reluctance for businesses and providers to act.[8][9][12]
- **UK market entry and compliance hurdles faced by foreign players:** As noted by a few US-based interviewees, market entry and compliance obstacles, coupled with complex public procurement processes, create challenges for foreign providers trying to access the UK market.[12]

Sources:

(1): Kharvi, Understanding the Impact of AI-Generated Deepfakes on Public Opinion, Political Discourse, and Personal Security in Social Media (https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10552098), 2024;

(2): Herbert Smith Freehills, Criminalising deepfakes – the UK's new offences following the Online Safety Act (https://www.herbertsmithfreehills.com/notes/tmt/2024-05/criminalising-deepfakes-the-uks-new-offences-following-the-online-safety-act), 2024;

(3): Rouse, AI-generated deepfakes: what does the law say? (https://rouse.com/insights/news/2024/ai-generated-deepfakes-what-does-the-law-say#:~:text=The%20European%20Union,procedures%20must%20also%20be%20provided.), 2024;

(4): Morrison Foerster, 2024 Year in Review: Navigating California's Landmark Deepfake Legislation (https://www.mofo.com/resources/insights/241211-2024-year-in-review-navigating-california-s-landmark-deepfake-legislation), 2024;

(5): Reed Smith, AI and publicity rights: The No Fakes Act strikes a chord (https://www.reedsmith.com/en/perspectives/2024/08/ai-and-publicity-rights-the-no-fakes-act-strikes-a-chord), 2024;

(6): Live Now Fox, What is the Take It Down Act? (https://www.livenowfox.com/news/take-down-act-what-to-know), 2025;

(7): Reuters, Spain to impose massive fines for not labelling AI-generated content (https://www.reuters.com/technology/artificial-intelligence/spain-impose-massive-fines-not-labelling-ai-generated-content-2025-03-11/), 2025;

(8): Tech Xplore, Spain seeks to criminalize AI-generated sexual images (https://techxplore.com/news/2025-03-spain-criminalize-ai-generated-sexual.html), 2025;

(9): Labuz. M., Deep fakes and the Artificial Intelligence Act—An important signal or a missed opportunity? (https://onlinelibrary.wiley.com/doi/epdf/10.1002/poi3.406), 2024;

(10): Howard Kennedy, The Regulatory Gap in Deepfake Technology (https://disputeresolution.howardkennedy.com/post/102jp32/the-regulatory-gap-in-deepfake-technology#:~:text=Despite%20growing%20awareness%20of%20the,AI%20and%20digital%20content%20creation.), 2025;

(11): GDPR Local, Deepfakes and the Future of AI Legislation: Overcoming the Ethical and Legal Challenges (https://gdprlocal.com/deepfakes-and-the-future-of-ai-legislation-overcoming-the-ethical-and-legal-challenges/), 2025.

(12): Interviews, 2025.

High costs, resource constraints, and low ROI perception are slowing deepfake detection innovations

Economic / Commercial trends and barriers
Rising enterprise losses calling for deepfake fraud prevention: Deepfakes are escalating financial risks across industries, enabling harms such as fraud, IP theft, brand damage, and executive impersonation, leading to significant economic losses for enterprises.(7) The inclusion of deepfake-enabled fraud is projected to drive total fraud losses in the U.S. to $40 billion by 2027, up from estimates that previously excluded AI-related threats.(1)

Provider willingness to partner: Industry players are increasingly forming partnerships with private and public entities to develop more effective solutions for addressing the diverse harms posed by deepfakes.(2)(3)(4)(5)(6) The UK Government is collaborating with industry and researchers to advance deepfake detection, supporting initiatives like the Deepfake Detection Challenge to address technological and regulatory challenges.(6)

Government funding on frontier AI safety driving foundational research: Government funding for frontier AI, such as the AI Security Institute's Systemic AI Safety Grants, supports research addressing AI-related risks, including emerging threats like deepfakes.(7)

Barriers

- **Customers perceiving low ROI in deepfake detection:** For customers, deepfake detection competes for resources against higher-priority areas such as product development, growth, and operational efficiency. As a result, investment in trust and safety often remains limited, slowing adoption and innovation in deepfake detection solutions.(7)
- **High costs and resource constraints:** Keeping pace with deepfake threats demands ongoing investment in AI research, computing power, and expertise. High development costs pose a barrier, especially for smaller deepfake detection providers, limiting market entry and innovation.(7)

Sources:

(1): Deloitte Center for Financial Services, Generative AI is expected to magnify the risk of deepfakes and other fraud in banking (https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html), 2024;

(2): Reality Defender, Partners (https://www.realitydefender.com/partners), 2025;

(3): Continuity Insights, New Partnership Will Advance Audio Deepfake Detection (https://continuityinsights.com/new-partnership-will-advance-audio-deepfake-detection/?utm_source=chatgpt.com), 2024;

(4): PR Newswire, Reality Defender Partners with TaskUs to Expand Deepfake Detection to Content Moderation Teams and Call Centers (https://www.prnewswire.com/news-releases/reality-defender-partners-with-taskus-to-expand-deepfake-detection-to-content-moderation-teams-and-call-centers-302312230.html?utm_source=chatgpt.com), 2024;

(5): Hive, Announcing Hive's Partnership with the Defense Innovation Unit (https://thehive.ai/blog/announcing-hives-partnership-with-the-defense-innovation-unit?utm_source=chatgpt.com), 2024;

(6): TechUK, UK Launches £200,000 Grants for Systemic AI Safety: Fostering Trust and Safety (https://www.techuk.org/resource/uk-launches-200-000-grants-for-systemic-ai-safety-fostering-trust-and-safety.html), 2024;

(7): Interviews, 2025.

Scepticism toward detection tools persists, likely due to a lack of confidence in their efficacy

Social trends and barriers

Increased public awareness of deepfake threats: There is rising public awareness and concern about deepfakes, particularly online CSEA, driven by high-profile cases and media coverage,[(1)(2)] potentially pushing political and policy attention toward addressing deepfakes.

Reputation: End user trust is a strong demand incentive for deepfake detection technology across a wide range of sectors. As new GenAI technology increases the number of crimes in various industries, both public and private institutions are more inclined to invest in deepfake detection services to maintain trust in their reputation.[(4)]

Barriers

- **End user scepticism toward detection tools:** As specified, users generally lack confidence in the ability of deepfake detection tools to distinguish real from manipulated content.(1) Users also appear to be overconfident in their own ability to identify deepfakes without technological assistance.(3) According to a survey conducted by the Alan Turing Institute, only 36.9% of the respondents trust AI-based technologies, compared to 54.4% for fact checking organisations and 46.3% for scientific experts.(1)(4)

Sources:

(1): The Alan Turing Institute, Behind the Deepfake: 8% Create; 90% Concerned (https://www.turing.ac.uk/sites/default/files/2024-07/behind_the_deepfake_full_publication.pdf), 2024;

(2): Kharvi, Understanding the Impact of AI-Generated Deepfakes on Public Opinion, Political Discourse, and Personal Security in Social Media (https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10552098), 2024;

(3): Business Wire, iProov Study Reveals Deepfake Blindspot: Only 0.1% of People Can Accurately Detect AI-Generated Deepfakes (https://www.businesswire.com/news/home/20250211131029/en/iProov-Study-Reveals-Deepfake-Blindspot-Only-0.1-of-People-Can-Accurately-Detect-AI-Generated-Deepfakes?utm_source=chatgpt.com), 2025;

(4): Interviews, 2025.

While GenAI advancements have spurred deepfake detection innovation, technical barriers exist

Technological trends and barriers

Increased GenAI-enabled threats: The rapid advancement of generative AI (GenAI) models - combined with their increasing realism, scalability, and accessibility - is accelerating deepfake-related threats, pushing the need for continuous innovation in detection technologies to keep pace.(1)(2)(3)(4)(5)(6)(10)

Development in AI advancing machine learning (ML) based deepfake detection: Ongoing innovations in AI-driven detection techniques are

enhancing the ability to identify and combat deepfakes. Improved machine learning models, multimodal detection approaches, and real-time analysis capabilities are driving more effective and scalable solutions, strengthening the deepfake detection market.(7)(10)

Barriers

- **Limited access to high-quality training datasets:** Most detection tools are trained on datasets that often lack real-world synthetic media, have limited coverage, and feature a disproportionate representation of deepfakes. Testing on these datasets can lead to inaccurate results. As noted by one interviewee, accuracy rates typically drop by 10-20% in real-world redeployment with representative datasets, compared to lab environments using training datasets.(8)(10)
- **Heterogeneity of testing datasets and metrics:** Inconsistencies in testing datasets and evaluation metrics result in low confidence in the efficacy of detection tools, while providers often claim high accuracy without independent validation.(9)(10)

Sources:

(1): techUK, Detecting deepfakes: A roadmap to UK resilience in the face of GenAI (https://www.techuk.org/resource/detecting-deepfakes-a-roadmap-to-uk-resilience-in-the-face-of-genai.html), 2024;

(2): Tech Target, How to prevent deepfakes in the era of generative AI (https://www.techtarget.com/searchsecurity/tip/How-to-prevent-deepfakes-in-the-era-of-generative-AI), 2024;

(3): Visua, The Deepfake Detection Arms Race (https://visua.com/the-deepfake-detection-arms-race?utm_source=chatgpt.com);

(4): DRCF, The Future of Synthetic Media (https://www.drcf.org.uk/publications/papers/the-future-of-synthetic-media/), 2024;

(5): Deepmedia, How Deepfakes Will Challenge Biometric Face Verification in 2025 (https://deepmedia.ai/blog/2025-biometric-face), 2025;

(6): Biometric Update, Deepfake ecosystem develops around apps, services as detection fights to keep pace (https://www.biometricupdate.com/202411/deepfake-ecosystem-develops-around-apps-services-as-detection-fights-to-keep-pace), 2024;

(7): GAO, Combating deepfakes (https://www.gao.gov/products/gao-24-107292), 2024,

(8): Yan et al., DeepfakeBench: A Comprehensive Benchmark of Deepfake Detection (https://arxiv.org/abs/2307.01426), 2023;

(9): UK Government, Innovating to detect deepfakes and protect the public (https://www.gov.uk/government/case-studies/innovating-to-detect-deepfakes-and-protect-the-public), 2025;

(10): Interviews, 2025.

3. Key Drivers Shaping the Future

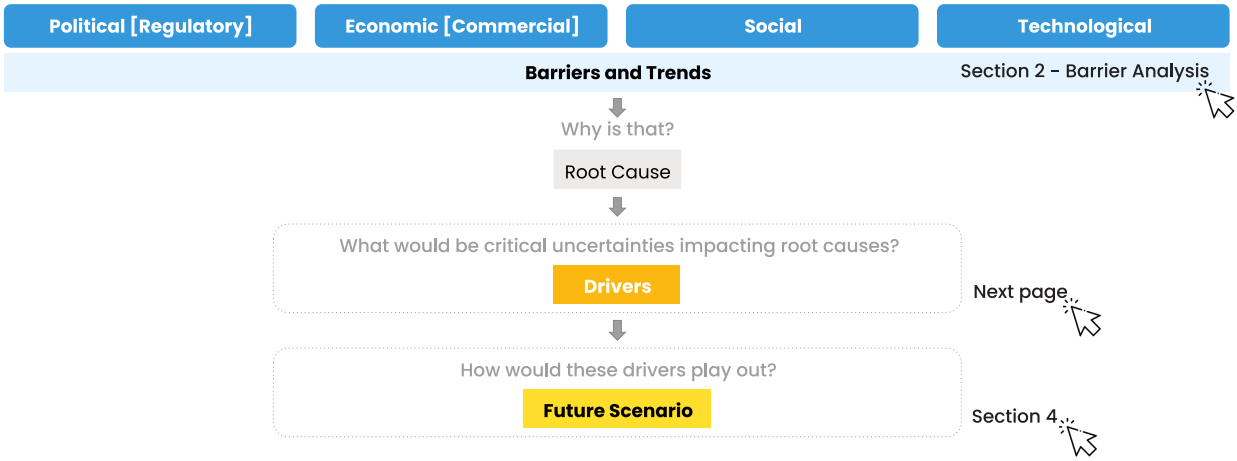
In this research, drivers are defined as critical uncertainties that could impact the deepfake detection market in the future.

Nine key drivers were identified, which interact to shape both demand and supply in the market. These include:

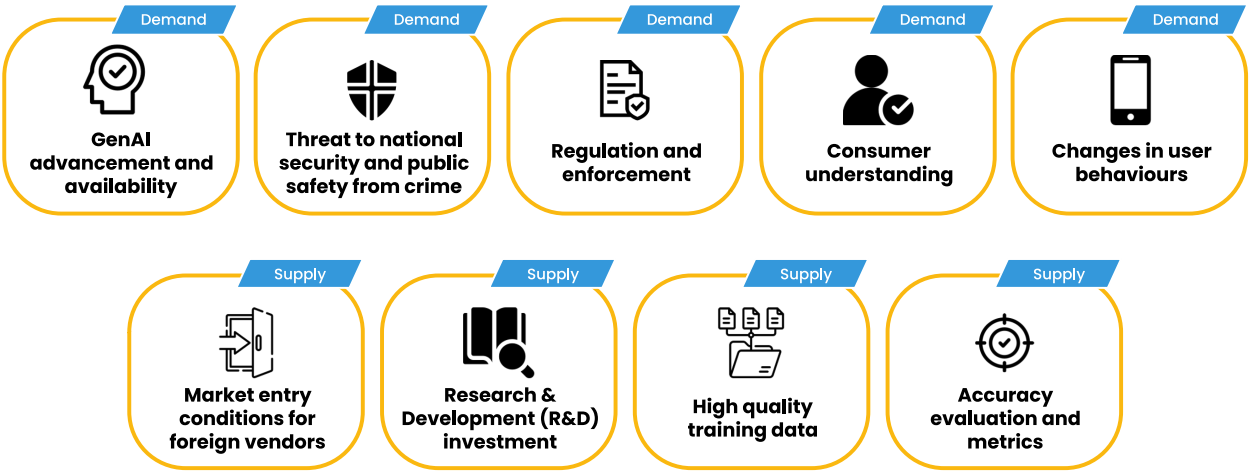
- Rapid advancement and availability of GenAI
- Threat to national security
- Regulation and enforcement mechanisms
- Customer understanding of deepfakes
- Changing user behaviours in response to deepfakes
- Market entry conditions for foreign vendors
- Available R&D investments
- Access to high-quality training data
- Standardised accuracy testing

Drivers are critical uncertainties that may have impacts on the deepfake detection market in the future

Approach to driver mapping
Synthesised from desk research of 86 sources and 14 expert interviews, primarily with deepfake detection providers



Nine key drivers interconnect to shape demand and supply in the deepfake detection market



GenAI advancements democratise deepfake creation and increase sophistication, calling for advanced detection

GenAI Advancement and Availability
As deepfake technology advances and related crimes become more frequent and financially significant, consumer awareness - among individuals and companies alike - has grown over the past few years.

This shift has altered their cost-benefit estimations, driving greater demand for deepfake detection tools and services.

Influences on the UK / International deepfake detection market

- There is a growing demand in sectors highly exposed to deepfake crimes, with increased awareness of GenAI-enabled threats.
- The ROI of deepfake detection investments has increased due to advancements in GenAI technology. These developments have led to more sophisticated crime techniques, posing increasingly significant risks - from financial and insurance fraud to CSEA and terrorism on online platforms.

Influences on sectors and industries

- New deepfake injection technologies have already been used in job interviews and fraud schemes, leading to multimillion-dollar losses and highlighting the urgent need for detection technologies.

> Further GenAI advancement would escalate threats and risks, likely **driving up demand** for deepfake detection technology, **incentivising investment**, and ultimately improving the quality of deepfake detection.

Democratisation of deepfake technology threatens national security and public safety, which would increase detection demand

Threat to National Security
Deepfakes hold the potential to diminish trust in public institutions and influence public opinion and elections by generating misleading content on sensitive topics and prominent figures. Public authorities are making early investments in detection technologies (e.g., via grant and challenge funds) and considering regulatory measures.

Influences on the UK / International deepfake detection market

- Growing national security concerns appear to be driving industry innovation in relevant technologies globally, with an increasing number of companies serving national security-related clients.

Influences on sectors and industries

- Governments and defence departments worldwide are increasingly concerned about the impact of deepfakes on misinformation, disinformation, and terrorism, showing early signs of investments in detection technology and research.

- Additionally, social media platforms are vulnerable, as deepfakes can be used to amplify terrorism threats and spread harmful, misleading content.

> **Governments** worldwide are increasingly aware of the **political risks** of **deepfakes**. This phenomenon is increasing the deepfake detection technology **demand** among authorities globally.

Effective regulation could stimulate demand for deepfake detection technology, thereby enhancing online safety

Regulation and Enforcement
Existing legislation, such as the EU AI Act, the UK's OSA, and various state-level regulations in the US, have laid the foundations in regulating against harmful content. If implemented effectively, these measures are likely to drive higher demand for detection technologies.

Influences on the UK / International deepfake detection market

- The UK's OSA already requires online platforms to implement specific content restrictions, which apply regardless of whether the content is real or AI-generated. Ofcom may want to consider whether AI specific guidance is needed in the future.

Influences on sectors and industries

- Online platforms regulated by the OSA, such as social media, gaming, and dating apps, would need to have effective content moderation solutions in place to combat CSEA and terrorism content, whether it is in the form of deepfakes or not.
- Financial institutions would need to re-assess their ROI if fraud-related losses were to become their responsibility, making it more likely that they allocate additional resources to detection tools and R&D.

> Stronger regulations could **shift the ROI calculation for deepfake detection customers** due to the fear of fines, therefore **driving demand** across sectors from finance to social platforms, especially where end users are more exposed to harmful synthetic content.

Demand for deepfake detection technology would also be driven by customer awareness and willingness to pay

Consumer Understanding of Deepfake Threats
Not every individual or industry exposed to deepfakes (whether through misinformation, disinformation, fraudulent intent, or public/private cyber theft) is aware of their multiple threats, nor are they familiar with detection tools. Increasing awareness of the potential risks posed by this type of content could drive greater demand for deepfake detection technologies.

Influences on the UK / International deepfake detection market

- Consumer understanding of these threats plays a significant role in driving the demand for detection technology, incentivising investment.
  - For example, two UK providers (Open Origins and Surf Security) secured early-stage funding in recent years, actively investing in deepfake detection capabilities driven by client needs.

Influences on sectors and industries

- A better understanding of the risks associated with deepfakes could drive increased demand for detection tools across industries. This is not only due to the rising sophistication of synthetic content but also the growing number of deepfake-related crimes, associated with harmful online activities and increasing financial losses.

> As deepfake-enabled crimes become more frequent, consumers' understanding of deepfake threats is likely to increase, **driving up the demand for both preventive and detection tools**, and boosting investment in technology.

End user engagement with technology would pressure businesses and government to act on deepfakes

Changes in User Behaviours
Currently, deepfakes can target a variety of devices and scenarios. For example, criminals can reach individuals via mobile phones or

organisations through hiring processes. This has broadened the use cases for deepfake detection and increased the need for novel detection techniques.

Influences on the UK / International deepfake detection market

- The evolving deepfake technology, which can impact everyday life, has prompted providers to invest in and continuously iterate their solutions in fraud prevention services, identity verification, and, more recently, video conferencing security.

Influences on sectors and industries

- As new devices (e.g., virtual reality devices) are targeted by deepfake creators seeking new ways to commit crimes (e.g., in the metaverse), the supply side would respond to these new manifestations of deepfake harms, potentially leading to increased investment in innovative detection techniques.

> The rise of deepfakes in professional and personal contexts would change how individuals interact with technology, making them **more cautious of potential deepfake threats**. This is likely to drive greater demand for detection solutions, **putting pressure on businesses and governments** to implement change.

Foreign providers have expressed interest in entering the

UK market but face compliance hurdles

Market Entry Conditions for Foreign Vendors
Public procurement processes in the UK can be challenging for foreign providers to navigate. For example, a few US providers have reported feeling restricted from participating in challenges and procurement processes compared to their UK counterparts as it requires them to join UK government frameworks.

There also other collaboration challenges due to data sharing and protection as US providers do not have GDPR.

Influences on the UK / International deepfake detection market

- Foreign companies have faced challenges when entering the UK market due to compliance hurdles. Additionally, when considering potential alliances, they find the legal structure for establishing joint ventures in the UK to be difficult, which further impacts their ability to compete and collaborate.

Influences on sectors and industries

- There are no sector-specific impacts related to this driver, as it pertains more to the general market dynamics rather than specific industries.

> Positioning the UK as an attractive market for foreign vendors by simplifying deepfake detection procurement and providing support would **enhance competition, promote knowledge sharing**, and **encourage innovation**.

Investment in deepfake detection R&D would drive the development of advanced solutions and economies of scale

Research & Development (R&D) Investment
Public and private investment in new detection technologies can enhance the supply of deepfake detection services, while simultaneously increasing consumer confidence in these tools, potentially boosting their demand.

Influences on the UK / International deepfake detection market

- Resource availability can limit economies of scale in AI-based technologies, as they often need significant computing power, resources, and specialised knowledge. More R&D investment in deepfake detection would help overcome these resource constraints and drive supply (both in quantity and diversity).
- Moreover, government-led R&D investment can signal market potential, providing greater certainty to both providers and private investors, encouraging market entry.

Influences on sectors and industries

- From the public investment perspective, consumption of deepfake detection solutions would increase as public investment in R&D improves access to this technology.

- Government signalling could spark interest from the private sector, driving demand and investment in this technology. This would not only boost confidence but also generate evidence and research on its performance.

> The increase in R&D on deepfake detection would boost the **supply and confidence** of the deepfake detection technology.

Availability of better training datasets can enhance

detection technology, building trust and increasing demand

High Quality Training Data
The lack of updated, representative datasets creates two main issues. First, it prevents deepfake detection tools from training on data generated with the latest creation technologies, making them incapable of detecting new nuances. Second, it undermines the reliability of testing results against such datasets, resulting in low trust and confidence in the detection tools.

Influences on the UK / International deepfake detection market

- The availability of better training datasets can influence the deepfake detection market by enhancing the accuracy and effectiveness of detection tools. High-quality, diverse, and up-to-date datasets allow detection providers to develop more robust algorithms, improving the reliability of their solutions across a wider range of deepfake types and use cases.
- This, in turn, increases the trust of both consumers and businesses in these technologies, driving demand.

Influences on sectors and industries

- With better training datasets, the market can also see increased innovation, as detection tools become more accurate and adaptable. This would lead to a broader range of applications, from law enforcement and security to media and entertainment, as well as the potential for faster adoption across industries.

> The availability of **high-quality training data** is crucial for **expanding the supply** of deepfake detection solutions and improving their quality through better training. As a result, **consumer trust would increase**, driving higher demand.

Standardised approach to accuracy evaluation can create fairer competition and enhance demand for products

Accuracy Evaluation and Metrics
There is no consensus on the accuracy metrics that providers use to promote their services, leading to consumer uncertainty and reluctance to pay for services they do not fully trust. Establishing a standardised way to evaluate accuracy to harmonise product results would enhance consumer trust and, in turn, drive increased spending on this type of technology.

Influences on the UK / international deepfake detection market

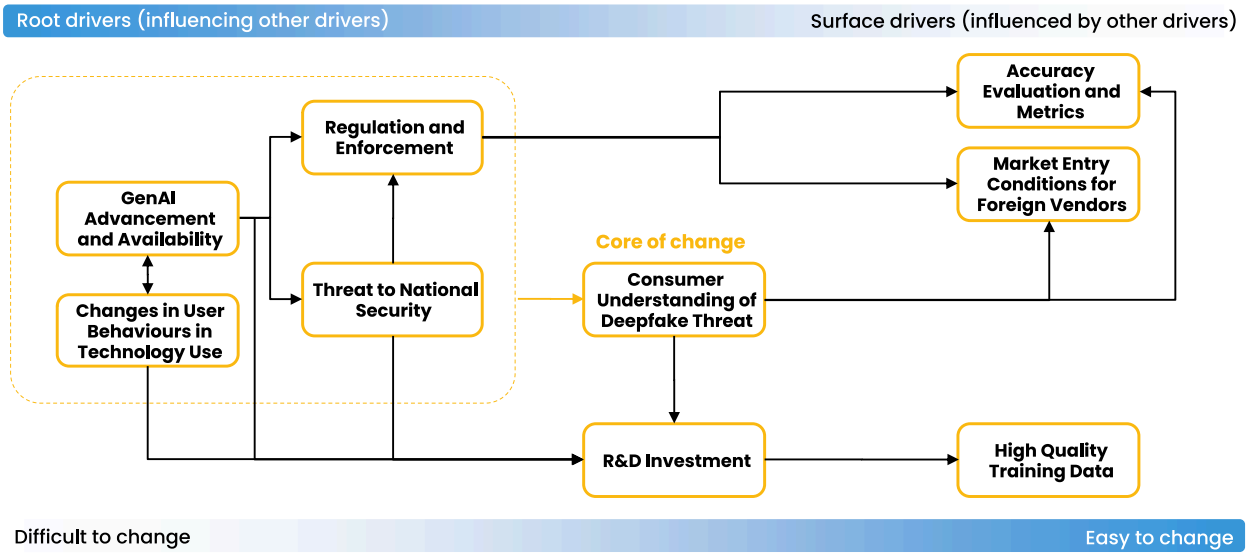
- The lack of clarity on accuracy metrics for deepfake solutions has led to consumer scepticism. Mitigating doubts about effectiveness, including the use of different datasets and metrics, is key to building trust.
- The nuance of real-world applications means that accuracy metrics will meaningfully differ depending on the use case in which deepfake detection tools are applied.

Influences on sectors and industries

- Clearer and standardised accuracy disclosure of detection tools would impact all industries vulnerable to deepfakes, as better information will reduce uncertainty and simplify procurement decisions.

> **A standardised approach** and guidance on accuracy evaluation **would benefit both consumers and providers**. It would align key considerations when conducting technical testing and interpreting the results, enabling meaningful comparison among different tools.

Mapping the interconnections among key drivers helps reveal where first mover interventions should target



Appendices

A1. Project Definitions & Glossary

Definitions

Barriers: Obstacles to entry and adoption in the deepfake detection market.

Deepfakes: Audio-visual content that has been generated or manipulated, and that misrepresents someone or something.(1)

Detection: Measures to identify and detect deepfake content after they have been created and shared.(1)

Drivers: Factors or influences shaping the future of the deepfake detection market. Drivers collectively impact the demand and supply of the detection market.

Embedding and Identification: Measures to attach and provide contextual information to content at the point of deepfake creation or editing to enable distinguishing them from real media.(1)

Glossary

AI: Artificial Intelligence

CSEA: Child Sexual Exploitation and Abuse

LLM: Large Language Model

ML: Machine Learning

OSA: Online Safety Act 2023

ROI: Return on investment

Sources: (1): Ofcom, Deepfake Defences: Mitigating the Harms of Deceptive Deepfakes, 2024

A2. Further Details on Methodology

Desk Research

Reviewed **86** relevant academic, industry, government and civil society reports, policy papers, and journal articles.

Obtained outputs:

- Key insights and trends identified from desk research were integrated with broader research findings throughout the report.

Expert Interviews

Conducted **14 one-to-one, semi-structured interviews**, subsequently compiling the insights into key barriers and drivers in the deepfake detection market. Targeted a mixed set of safety tech providers prioritised by geography and organisational size.

Obtained outputs:

- Identified challenges and drivers in the deepfake detection industry.
- Key findings highlighted the components and motivations on the demand side, as well as pain points on the supply side. These insights informed this report.

Futures Workshop

Developed a **futures workshop** with 13 participants from government, regulators, and deepfake detection providers to validate drivers, define the current and desired deepfake detection market scenarios.

Obtained outputs:

- Defined the depth of the drivers and the influence of authorities on them while establishing policy implementation priorities.
- Key insights into the market scenarios from both demand and supply sides, highlighting possible recommendations informed in this report.

A2. Further Details on Methodology - Stakeholder Engagement

Below is a list of organisations engaged to gather insights through the course of this research:

Deepfake Detection & Safety Tech Providers

- Deep Media
- GetReal Labs
- Hive
- Naimuri
- Videntifier
- Open Origins
- Privately
- Reality Defender
- Surf Security
- Yoti

Industry Associations

- Age Verification Providers Association (AVPA)
- Online Safety Tech Industry Association (OSTIA)

Non-profits

- Internet Watch Foundation

Government & Regulators

- Department for Science, Innovation and Technology (DSIT)
- Office of Communications (Ofcom)
- National Crime Agency (NCA)

A2. Further Details on Methodology - Third-party Deepfake Detection Provider Mapping

Identify	Filter	Cross-check	Aggregate	Enrich
Conducted thorough research across reports, articles, industry news, and relevant websites to **compile an initial** list of global deepfake detection providers.	**Manually reviewed lists** of global Safety Tech providers with potential deepfake relevance from various sources, including Crunchbase. **Applied inclusion criteria** (i.e., active deepfake detection solutions, commercialisation, excluding irrelevant providers) to identify and scope companies **actively involved in deepfake detection.**	**Compared** the initial deepfake detection list with new filtered lists to **identify any missing companies** and incorporated them as needed.	**Merged and standardised the lists** into a structured dataset, categorising companies by key attributes such as geography, market focus (dedicated vs. diversified), and detection capabilities (images, audio, video, text).	**Collected additiona data** to re market insights fr **online aggregat** (e.g., Crunchba **available web data** and **manu review** of company websitesc • Compan size • Funding • Partnersh • Technol capability • Use cas • Custome segments • Accurac claims

A2. Further Details on Methodology - Futures Workshop

The workshop was **not** about predicting a certain future. It was about exploring a range of plausible futures, strengthening future facing policies

and strategies, and encouraging collaboration and trust among public sector stakeholders.

Before the Workshop

A list of 9 drivers affecting the future was identified based on desk research and expert interviews.

During the Workshop

Part 1 - Scenario Framework

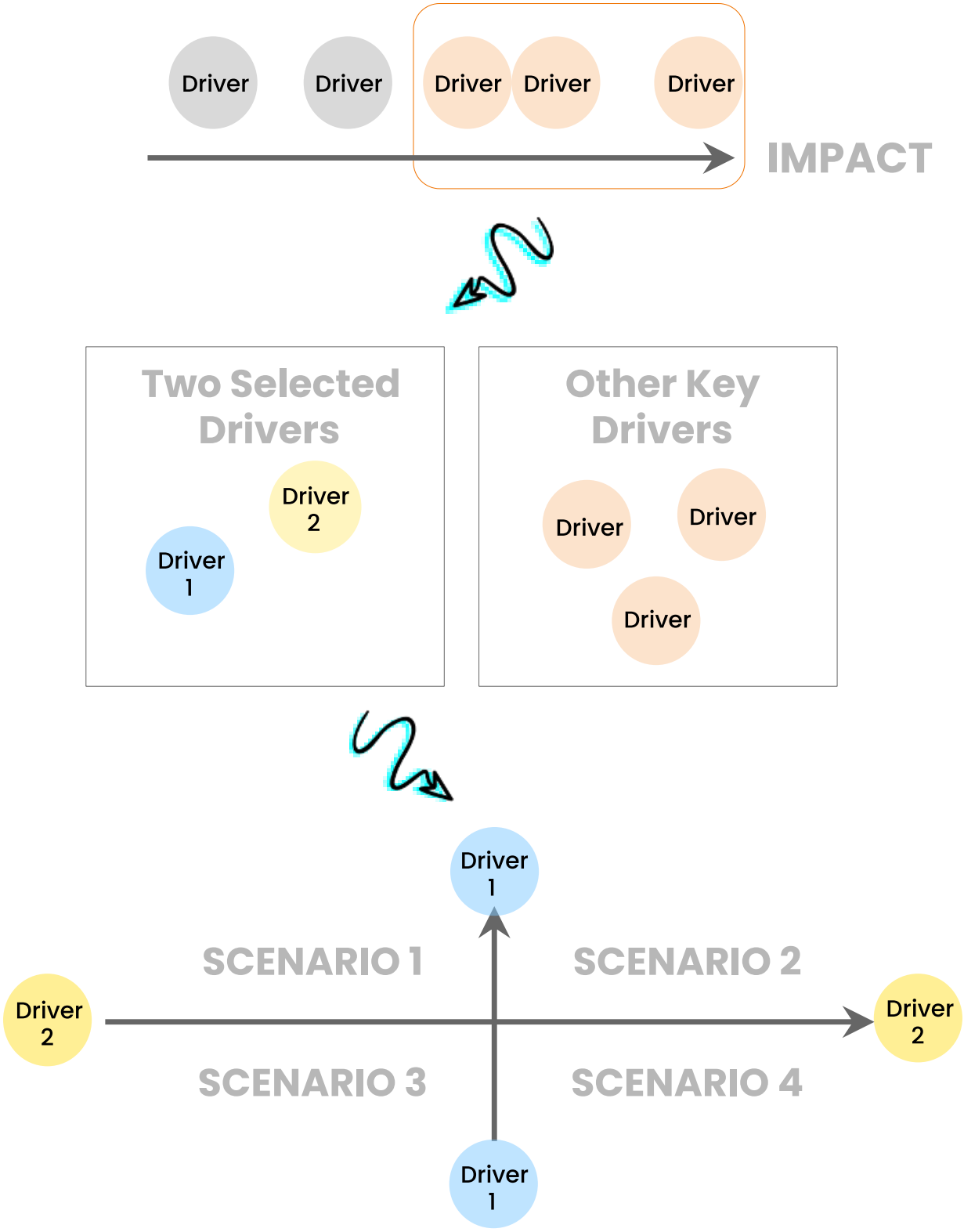
- Mapped drives an impact scale individually.
- Identified the most impactful drivers through facilitated group discussion.
- Selected 2 drivers that form the axes of the 2x2 scenarios matrix to structure the Part 2 discussion.

> Which drivers should we pay attention to?
>
> Which drivers can the government feasibly influence?

Part 2 - Scenario Building

- Develop and identified current and future scenarios.
- Discuss the interventions or strategies that can help achieve or avoid certain scenario(s).

> What are the main players in this scenario?
>
> How would each key driver play out in the scenarios?
>
> Is it a scenario we want or not?



A2. Further Details on Methodology - Limitations

Several limitations should be noted alongside the research findings

Provider mapping:

The provider mapping built upon previous research into the safety tech ecosystem, including the DSIT's UK Safety Tech Sectoral Analysis. Additional providers were identified via available web data, online aggregators (e.g., Crunchbase), and manual keyword searches and reviews.

The focus of this project was on qualitative analysis over quantitative analysis, **resulting in gaps in the economic data and the quantification** of the size, growth and benefits of deepfake detection.

The provider mapping aimed to establish a foundational baseline for the sector's future development. **As such, it should not be viewed a one-time effort or an exhaustive list of providers.**

The mapping focused on third-party providers, while recognising that several larger tech companies like Microsoft, Intel, Meta, and Google DeepMind develop in-house detection solutions (as shown in slide 25).

Stakeholder engagement:

The stakeholder engagement for this research primarily focused on expert interviews with industry providers who have established deepfake detection capabilities or are currently exploring these capabilities, **focusing on the supply side of the market.**

Engagement with the demand side was relatively limited, with a small group from the public sector (e.g., DSIT, NCA, Ofcom) participating in a workshop or interviews. As a result, there is a gap in the direct engagement with other key stakeholders and customers of deepfake detection from both government and industry sectors (including GenAI developers and application providers). This presents potential for future engagement.

A3. Bibliography

1. Abbas & Taeihagh, Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence (https://www.sciencedirect.com/science/article/pii/S0957417424011266#s0045), 2024
2. Biometric Update, Deepfake ecosystem develops around apps, services as detection fights to keep pace (https://www.biometricupdate.com/202411/deepfake-ecosystem-develops-around-apps-services-as-detection-fights-to-keep-pace), 2024
3. Biometric Update, VCs like Clarity's approach to deepfake detection, but there are alternatives (https://www.biometricupdate.com/202402/vcs-like-claritys-approach-to-deepfake-detection-but-there-are-alternatives), 2024

4. Business Wire, iProov Study Reveals Deepfake Blindspot: Only 0.1% of People Can Accurately Detect AI-Generated Deepfakes (https://www.businesswire.com/news/home/20250211131029/en/iProov-Study-Reveals-Deepfake-Blindspot-Only-0.1-of-People-Can-Accurately-Detect-AI-Generated-Deepfakes?utm_source=chatgpt.com), 2025
5. Continuity Insights, New Partnership Will Advance Audio Deepfake Detection (https://continuityinsights.com/new-partnership-will-advance-audio-deepfake-detection/?utm_source=chatgpt.com), 2024
6. DASH Lab, FakeAVCeleb: A Novel Audio-video Multimodal Deepfake Dataset [GitHub repository] (https://github.com/DASH-Lab/FakeAVCeleb), 2025
7. Diel, A., et al., Human performance in detecting deepfakes (https://www.sciencedirect.com/science/article/pii/S2451958824001714#sec5), 2024
8. Deepmedia.ai, Generator Attribution in GenAI Images - “Laboratory” Benchmark and Accuracy Report (https://deepmedia.ai/research/generator-attribution-in-genai-images-laboratory-benchmark-and-accuracy-report), 2024
9. Deepmedia.ai, How Deepfakes Will Challenge Government Agencies in 2025 (https://deepmedia.ai/research/generator-attribution-in-genai-images-laboratory-benchmark-and-accuracy-report), 2025
10. Deepmedia.ai, How Deepfakes Will Challenge Biometric Face Verification in 2025 (https://deepmedia.ai/blog/2025-biometric-face), 2025
11. DeepTrust, DeepTrust and Breacher AI Partnership Announcement (https://www.deeptrust.ai/blog/deeptrust-and-breacher-ai-partnership-announcement-2/), 2024
12. DeepTrust, DeepTrust and Musicfy Partnership Announcement (https://www.deeptrust.ai/blog/deeptrust-and-musicfy-partnership-announcement/), 2024
13. Deepware, Terms of Services (https://deepware.ai/terms-of-services/)
14. Deloitte Center for Financial Services, Generative AI is expected to magnify the risk of deepfakes and other fraud in banking (https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html), 2024
15. Dolhansky et al, The Deepfake Detection Challenge (DFDC) Dataset (https://arxiv.org/abs/2006.07397), 2020
16. DSIT, The UK Safety Tech Sector: 2024 Analysis (https://assets.publishing.service.gov.uk/media/6707a9f030536cb927482f69/uk_safety_tech_sector_2024_analysis.pdf), 2024
17. DRCF, The Future of Synthetic Media (https://www.drcf.org.uk/siteassets/drcf/pdf-files/the-future-of-synthetic-media.pdf?v=385978), 2024

18. Esri, How Compute Accuracy For Object Detection works (https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/how-compute-accuracy-for-object-detection-works.htm#:~:text=F1%20score%E2%80%94The%20F1%20score,where%201%20means%20highest%20accuracy.&text=Precision%2Drecall%20curve%E2%80%94This%20is,of%20an%20object%20detection%20model.)
19. Expert Insights, The Top 8 Deepfake Detection Solutions (https://expertinsights.com/insights/the-top-deepfake-detection-solutions/), 2024
20. Forbes, The Year Of The Deepfake: Combating Digital Deception In 2024 And Beyond (https://www.forbes.com/councils/forbestechcouncil/2024/12/30/the-year-of-the-deepfake-combating-digital-deception-in-2024-and-beyond/), 2024
21. GAO, Combating deepfakes (https://www.gao.gov/products/gao-24-107292), 2024
22. Gartner, Gartner Predicts 30% of Enterprises Will Consider Identity Verification and Authentication Solutions Unreliable in Isolation Due to AI-Generated Deepfakes by 2026 (https://www.gartner.com/en/newsroom/press-releases/2024-02-01-gartner-predicts-30-percent-of-enterprises-will-consider-identity-verification-and-authentication-solutions-unreliable-in-isolation-due-to-deepfakes-by-2026), 2024
23. GDPR Local, Deepfakes and the Future of AI Legislation: Overcoming the Ethical and Legal Challenges (https://gdprlocal.com/deepfakes-and-the-future-of-ai-legislation-overcoming-the-ethical-and-legal-challenges/), 2025
24. Global View Research, Fraud Detection And Prevention Market Size, Share & Trends Analysis Report (https://www.grandviewresearch.com/industry-analysis/fraud-detection-prevention-market#:~:text=Market%20Size%20%26%20Trends,18.7%25%20from%202025%20to%202030.), 2024
25. Google Deepmind, SynthID (https://deepmind.google/technologies/synthid/)
26. Groh et al, Human detection of political speech deepfakes across transcripts, audio, and video (https://www.media.mit.edu/publications/human-detection-of-political-deepfakes-across-transcripts-audio-and-video/), 2024
27. Heidari et al, Deepfake detection using deep learning methods: A systematic and comprehensive review (https://wires.onlinelibrary.wiley.com/doi/epdf/10.1002/widm.1520), 2022
28. Herbert Smith Freehills, Criminalising deepfakes – the UK's new offences following the Online Safety Act (https://www.herbertsmithfreehills.com/notes/tmt/2024-05/criminalising-deepfakes-the-uks-new-offences-following-the-online-safety-act), 2024
29. Hive, Announcing Hive's Partnership with the Defense Innovation Unit (https://thehive.ai/blog/announcing-hives-partnership-with-the-defense-innovation-unit?utm_source=chatgpt.com), 2024

30. Home Office, Launching the Deepfake Detection Challenge: a collaborative effort against digital deception (https://ace.blog.gov.uk/2024/05/10/launching-the-deepfake-detection-challenge-a-collaborative-effort-against-digital-deception/), 2024

31. Howard Kennedy, The Regulatory Gap in Deepfake Technology (https://disputeresolution.howardkennedy.com/post/102jp32/the-regulatory-gap-in-deepfake-technology#:~:text=Despite%20growing%20awareness%20of%20the,AI%20and%20digital%20content%20creation.), 2025

32. Innovatrix, Liveness Detection (https://www.innovatrics.com/glossary/liveness-detection/)

33. Intel, Real-time FakeCatcher for Deepfake Detection (https://www.intel.com/content/www/us/en/research/trusted-media-deepfake-detection.html)

34. Interviews, 2025

35. Jumio, Jumio Adds iProov's Award-Winning Liveness Detection to its KYX Platform (https://www.jumio.com/about/press-releases/iproov-liveness-detection-kyx/), 2021

36. Kaur et al, Deepfake video detection: challenges and opportunities (https://link.springer.com/article/10.1007/s10462-024-10810-6), 2024

37. Kharvi, P., Understanding the Impact of AI-Generated Deepfakes on Public Opinion, Political Discourse, and Personal Security in Social Media (https://www.semanticscholar.org/paper/Understanding-the-Impact-of-AI-Generated-Deepfakes-Kharvi/d3b9580d173e81643e15bfa01e7ef9b37acd4127), 2024

38. Knowledge Futures, Trusting Video in the Age of Generative AI (https://commonplace.knowledgefutures.org/pub/9q6dd6lg/release/2), 2023

39. Labuz. M., Deep fakes and the Artificial Intelligence Act—An important signal or a missed opportunity? (https://onlinelibrary.wiley.com/doi/epdf/10.1002/poi3.406), 2024

40. Le et al, SoK: Facial Deepfake Detectors (https://arxiv.org/pdf/2401.04364), 2024

41. MarketsandMarkets, The Rise of Deepfake AI Market: A $5,134 million Industry Dominated by Tech Giants (https://www.globenewswire.com/news-release/2024/07/17/2914676/0/en/The-Rise-of-Deepfake-AI-Market-A-5-134-million-Industry-Dominated-by-Tech-Giants-Synthesia-Reface-Sentinel-AI-Pindrop-BioID-MarketsandMarkets.html), 2024

42. Meta, Video Seal - Meta FAIR AI Demos (https://aidemos.meta.com/videoseal)

43. Microsoft, New Steps to Combat Disinformation (https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/?utm_source=chatgpt.com), 2020

44. Morrison Foerster, 2024 Year in Review: Navigating California's Landmark Deepfake Legislation (https://www.mofo.com/resources/insights/241211-2024-year-in-review-navigating-california-s-landmark-deepfake-legislation), 2024
45. Monteiro, Wanzeller, and Caldeira, Performance Analysis on DeepFake Detection (https://ibimapublishing.com/articles/CIBIMA/2024/457767/457767.pdf), 2024
46. Mutka et al, An Investigation of the Effectiveness of Deepfake Models and Tools (https://www.mdpi.com/2224-2708/12/4/61), 2023
47. Norwest Venture Partners, Rising Demand for Deepfake Security Solutions: How AI Is Enabling a New Frontier of Phishing Attacks (https://www.nvp.com/blog/deepfake-detection-solutions-ai-new-frontier-phishing-attacks/), 2024
48. Ofcom, Deepfake Defences: Mitigating the Harms of Deceptive Deepfakes (https://www.ofcom.org.uk/siteassets/resources/documents/consultations/discussion-papers/deepfake-defences/deepfake-defences.pdf?v=370754), 2024
49. PA Consulting, Deepfakes: a human challenge (https://www.weprotect.org/wp-content/uploads/Deepfakes_A-Human-Challenge_PA-Report_v3.pdf), 2024
50. Patil et al, Deepfake Detection using Biological Features: A Survey (https://arxiv.org/pdf/2301.05819), 2023
51. Payments Cards & Mobile, Fighting deepfakes has become a billion-pound industry in the UK alone (https://www.paymentscardsandmobile.com/fighting-deepfakes-has-become-a-billion-pound-industry-in-the-uk-alone/#:~:text=862,become%20a%20billion%20pound%20industry.), 2024
52. Perspective Economics, The UK Safety Tech Sector: 2023 Analysis (https://assets.publishing.service.gov.uk/media/647755195f7bb7000c7fa291/uk_safety_tech_analysis_2023.pdf), 2023
53. Pu et al, Deepfake Text Detection: Limitations and Opportunities (https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10179387), 2023
54. PR Newswire, Reality Defender Partners with TaskUs to Expand Deepfake Detection to Content Moderation Teams and Call Centers (https://www.prnewswire.com/news-releases/reality-defender-partners-with-taskus-to-expand-deepfake-detection-to-content-moderation-teams-and-call-centers-302312230.html?utm_source=chatgpt.com), 2024
55. Public Citizen, Tracker: State Legislation on Deepfakes in Elections (https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/), 2025
56. Reality Defender, AVFF: Audio-Visual Feature Fusion for Video Deepfake Detection (https://www.realitydefender.com/blog/audio-visual-feature-fusion-for-video-deepfake-detection), 2024

57. Reality Defender, Partners (https://www.realitydefender.com/partners), 2025

58. Reality Defender, Strengthening Digital Trust: How ElevenLabs Helps Reality Defender Advance Voice Deepfake Detection (https://www.realitydefender.com/blog/elevenlabs-reality-defender-voice-deepfake-detection), 2025

59. Reed Smith, AI and publicity rights: The No Fakes Act strikes a chord (https://www.reedsmith.com/en/perspectives/2024/08/ai-and-publicity-rights-the-no-fakes-act-strikes-a-chord), 2024

60. Resemble.AI, Deepfake Threats: An Analysis of Enterprise Risk and Mitigation in 2024 (https://www.resemble.ai/wp-content/uploads/2024/11/Deepfake-Threats-2024.pdf), 2024

61. Responsible AI, A Look at Global Deepfake Regulation Approaches (https://www.responsible.ai/a-look-at-global-deepfake-regulation-approaches/), 2023

62. Reuters Institute, Spotting the deepfakes in this year of elections: how AI detection tools work and where they fail (https://reutersinstitute.politics.ox.ac.uk/news/spotting-deepfakes-year-elections-how-ai-detection-tools-work-and-where-they-fail), 2024

63. Rouse, AI-generated deepfakes: what does the law say? (https://rouse.com/insights/news/2024/ai-generated-deepfakes-what-does-the-law-say#:~:text=The%20European%20Union,procedures%20must%20also%20be%20provided.), 2024

64. Security Brief, SURF unveils AI tool to combat growing deepfake threats (https://securitybrief.co.uk/story/surf-unveils-ai-tool-to-combat-growing-deepfake-threats), 2024

65. Sensity, The State of Deepfakes 2024 (https://5865987.fs1.hubspotusercontent-na1.net/hubfs/5865987/SODF%202024.pdf), 2024

66. Sharma, Garg & Caudron, A systematic literature review on deepfake detection techniques (https://link.springer.com/article/10.1007/s11042-024-19906-1), 2024

67. Slashdot, Best Deepfake Detection Software in the UK (https://slashdot.org/software/deepfake-detection/), 2024

68. Synesthesia.io, The Future of (Synthetic) Media (https://www.synthesia.io/post/the-future-of-synthetic-media#what-are-synthetic-and-non-synthetic-media), 2024

69. Tan et al, Rethinking the Up-Sampling Operations in CNN-based Generative Network for Generalizable Deepfake Detection (https://openaccess.thecvf.com/content/CVPR2024/papers/Tan_Rethinking_the_Up-Sampling_Operations_in_CNN-based_Generative_Network_for_Generalizable_CVPR_2024_paper.pdf), 2023

70. TechRound, Top 10 Deep Fake Detection Tools (https://techround.co.uk/tech/deep-fake-detection-tools/), 2024

71. TechUK, Detecting deepfakes: A roadmap to UK resilience in the face of GenAI (https://www.techuk.org/resource/detecting-deepfakes-a-roadmap-to-uk-resilience-in-the-face-of-genai.html), 2024

72. TechUK, Deepfakes and Synthetic Media: What are they and how are techUK members taking steps to tackle misinformation and fraud (https://www.techuk.org/resource/synthetic-media-what-are-they-and-how-are-techuk-members-taking-steps-to-tackle-misinformation-and-fraud.html), 2023

73. TechUK, Discussing Deepfakes: The opportunities and challenges of synthetic media technology (https://www.techuk.org/resource/discussing-deepfakes-how-could-we-respond-to-the-threat-of-new-deepfake-technologies-and-seize-the-potential-of-synthetic-media-technologies.html), 2024

74. TechUK, UK Launches £200,000 Grants for Systemic AI Safety: Fostering Trust and Safety (https://www.techuk.org/resource/uk-launches-200-000-grants-for-systemic-ai-safety-fostering-trust-and-safety.html), 2024

75. The Alan Turing Institute, Behind the Deepfake: 8% Create; 90% Concerned (https://www.turing.ac.uk/sites/default/files/2024-07/behind_the_deepfake_full_publication.pdf), 2024

76. The Alan Turing Institute, What are deepfakes and how can we detect them? (https://www.turing.ac.uk/blog/what-are-deepfakes-and-how-can-we-detect-them), 2024

77. The Alan Turing Institute, 9 in 10 concerned about deepfakes affecting election results (https://www.turing.ac.uk/news/9-10-concerned-about-deepfakes-affecting-election-results), 2024

78. The Centre for Policy Studies, Facing Fakes (https://cps.org.uk/research/facing-fakes/), 2024

79. The Constitution Society, The Democratic Challenges of Deepfake Technology (https://consoc.org.uk/democratic-challenges-of-deepfake-tech/), 2024

80. University of Maryland, Fighting Deepfakes, Shallowfakes and Media Manipulation (https://cmns.umd.edu/news-events/news/Nirupam-Roy-fighting-deepfakes-shallowfakes-and-media-manipulation), 2024

81. UK Government, Innovating to detect deepfakes and protect the public (https://www.gov.uk/government/case-studies/innovating-to-detect-deepfakes-and-protect-the-public), 2025

82. UK Government, Introducing the AI Safety Institute (https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute#box-1), 2024

83. Visua, The Deepfake Detection Arms Race (https://visua.com/the-deepfake-detection-arms-race?utm_source=chatgpt.com)

84. Yan et al, DeepfakeBench: A Comprehensive Benchmark of Deepfake Detection (https://arxiv.org/abs/2307.01426), 2023

85. Yi et al, Audio Deepfake Detection: A Survey (https://arxiv.org/abs/2308.14970), 2023

86. Zhu et al (Reality Defender), Learn from Real: Reality Defender's Submission to ASVspoof5 Challenge (https://arxiv.org/abs/2410.07379), 2024

About PUBLIC

This work was prepared by a specialist digital advisory firm, PUBLIC, who have been supporting the department with this analysis. PUBLIC have designed and delivered a number of UK Government programmes designed to support high-growth technology sectors, including with the Department for Science, Innovation and Technology, the Ministry of Housing, Communities and Local Government, the Ministry of Justice, the NHS, and the Geospatial Commission.

This work is not official government policy.