



Towards a Safer Nation: The United States 'Safety Tech' Market

RESEARCH COMMISSIONED BY PALADIN CAPITAL GROUP



Contents

Foreword	2	Products and Services	25
Executive Summary	4	Location	28
Introduction	7	Number of Safety Tech Companies: Time Series	30
Defining Safety Tech	8	Estimated Employment	32
Methodology	10	Investment in US Safety Tech	32
Why does Safety Tech matter?	12	Market Trends & Opportunities	41
Position Paper: The Cyber Blue Line	15	International Comparison	42
United States: Safety Tech Policy	16	Appendix	46
Safety Tech Taxonomy	19		
The US Safety Tech Market	23		

Foreword

Safety and Security in the Cyber Age demands a full array of new solutions and innovations anchored deeply in technology and human experiences. Adoption of new digital platforms has led to new threats and vulnerabilities for users in general, and specifically against children and those who are vulnerable, as well as threats to the critical infrastructure that we rely on each and every day. As much as we need to protect the resiliency of data and critical infrastructure, we must prioritize solutions that protect the safety, security and resiliency of people interacting in a digital world.

Paladin, since its founding 20 years ago, has promoted a deep culture that sought to deliver technology solutions to meet the urgent needs of our culture, business, government and society. Needs, that if not attended to, would cause our way of life to deteriorate and fail.

Hence Paladin's intense focus on Safety Tech. Adversaries now wage conflicts and sow public distrust at the human level through online harassment, misinformation and deception. They attack with harmful conduct, content and contact.

Such cyberwarfare stands to be near constant, generating crises for individuals and groups. These crises divide and degrade the well-being of individuals, families, communities and democratic societies. Paladin believes now is the time to intensify investing in Safety Tech solutions

for human-level protection and security in cyberspace. We see Safety Tech as an innovative ecosystem, an emerging global industry to align closely with Paladin's mission, investment thesis, and expertise.

This report represents Paladin's view on the need for and the state of Safety Tech in the United States. It highlights the case for public policy action and private investment in Safety Tech. Paladin Capital Group welcomes public and private partnerships to bring innovative solutions to this critically important sector of our digitally connected world. We encourage stakeholders to connect with our team to build the partnerships critical for success in this new industry. We thank Dr. Mary Aiken for her decades of work on this topic and her ongoing support and guidance to Paladin.



MICHAEL STEED
FOUNDER AND MANAGING PARTNER
PALADIN CAPITAL GROUP

Foreword

A new sector, the online safety technology or 'Safety Tech' sector, which complements the existing cybersecurity industry is gaining prominence. This research report has found evidence of an emerging and thriving US Safety Tech sector that aims to deliver solutions to facilitate safer online experiences and protect people from psychological risks, criminal dangers and online harm. Importantly, Safety Tech innovations also have the capacity to protect people from the corrosive effects of misinformation, online harassment, discrimination, and extremism which increasingly threaten democracy and civil society.

What is the difference between cybersecurity and cyber safety? Binary; cybersecurity primarily focuses on protecting data, systems and networks; cyber safety or Safety Tech focuses on protecting people.

It is critical that data, information, systems and networks are protected from cyber-attacks and are robust, resilient and secure. However, it is equally critical that the people who operate and use these systems are psychologically robust, resilient, safe and secure. Therefore, it is the combination of cybersecurity and Safety Tech that will deliver optimum protection.

The pandemic has changed how we live, accelerating the global digital pivot and exponentially increasing associated risks. Safety Tech aims to ensure that people are afforded the utmost protection when interacting with technology and that the levels of assurance we expect in the real world are matched in cyber-contexts. The construct of 'Online Harm'¹ is now becoming well-established worldwide, consisting of a spectrum of harm mediated by technology,

ranging from harassment and targeted abuse to misinformation and hate speech. Definitions of harm, the role of technology and governance responsibilities are increasingly the subject of legal debate and discussion in the US, with particular consideration regarding first amendment protections and freedom of speech concerns.

Tackling online harms will require input from government, non-profits, regulators and the private sector. Providing investment, support and enhancing awareness of Safety Tech system, platform and endpoint solutions is imperative to deliver technology solutions to technology-facilitated online harms. Effectively, Safety Tech is about cyber leadership, it is about values, reputation, and sustainable businesses in a global context. All of this can help to shape the kinds of society that we want to live in and enjoy.

The Safety Tech sector offers new possibilities regarding impact investing, that is, strategic investment opportunities that could enhance the environment of cyberspace. Arguably, a new evolution in environmental, social, and governance (ESG) investment criteria, importantly, helping to create safer and more secure cyber societies.



PROFESSOR MARY AIKEN
PALADIN STRATEGIC
ADVISORY GROUP

¹ UK Government (2020) 'Online Harms White Paper' Available at: <https://www.gov.uk/government/consultations/online-harms-white-paper>

Executive Summary

KEY FINDINGS AND REFLECTIONS: THE START OF A VERY IMPORTANT CONVERSATION...

This research seeks to identify providers of Safety Tech products or services, with a clear presence in the United States market, that are active and undertake commercial activity.

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

This research sets out a baseline within the United States for a high-growth market that has the potential to benefit online users globally and keep people safer online.

- We have identified more than **160 dedicated SafetyTech businesses** operating in the US market. We recognize there will be many more within aligned industries that will have considerable potential to grow the sector further.
- We estimate there is a community of more than **8,800 Safety Tech professionals** (within these businesses) in the US.

- The sector has been growing rapidly, given the rise in online harms such as hateful content and misinformation. This has increased the need by industry, government and individuals to explore solutions to help address and counter these harms. As a result, there has been **80% growth** in the number of active Safety Tech firms in the United States since 2016.
- There is a wide range of factors driving **demand for Safety Tech solutions**, including an increased risk and threat landscape with respect to online harms, and the recognition that user safety is a key component of supporting positive online experiences.
- Investors have started to recognise the need for these solutions. **Over \$1bn has been raised in external investment** by these firms in recent years, with 2020 being a record year for investment in Safety Tech (\$228m) in the US.
- This aligns strongly to **impact investing, Responsible Investing (RI); Environmental, Social and Governance (ESG);** and ethical investment principles. We believe that further investment in this space will enable the development and expansion of ethical and effective techniques to address a range of online harms.

We hope that this research promotes further debate and conversation about the role of the Safety Tech market in addressing online harms. Input from industry, government agencies, law enforcement, policymakers, academia, educational institutions, social media companies, NGO's and all stakeholders can act as a catalyst to grow and support these businesses in making the Internet a safer and more secure place, and in doing so work towards building a safer nation.



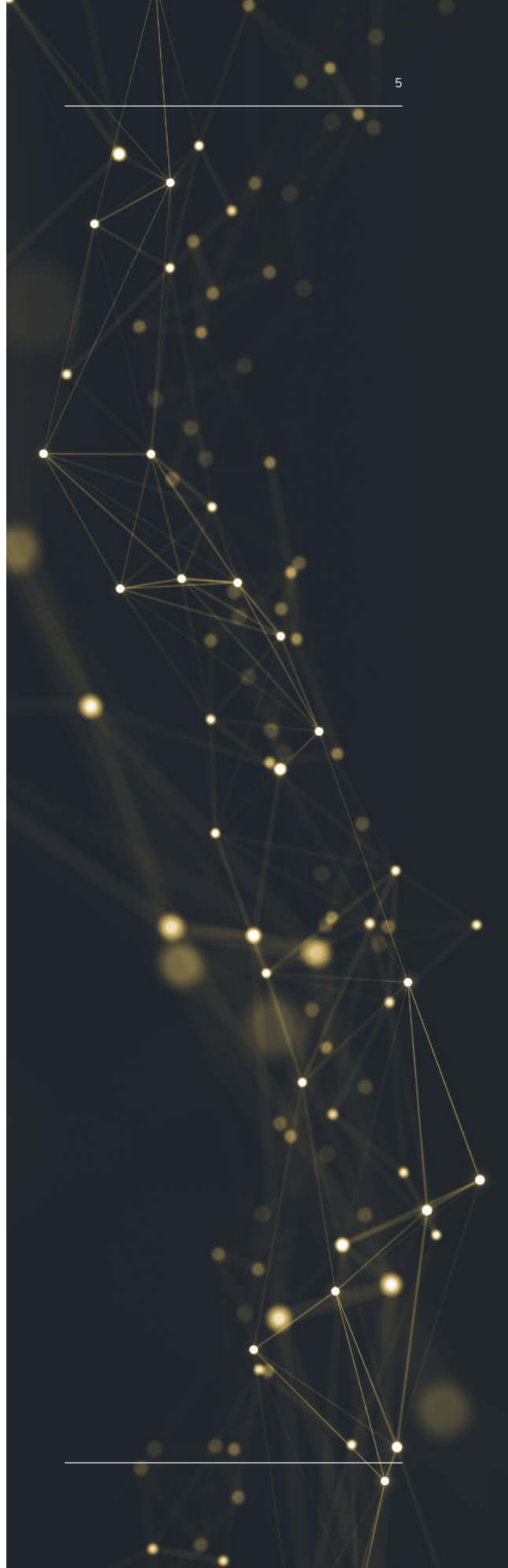
TOWARDS A SAFER NATION

This moment calls upon us to lean forward, not shrink back – to boldly engage the world to keep Americans safe, prosperous, and free. It requires a new and broader understanding of national security, one that recognizes that our role in the world depends upon our strength and vitality here at home. It demands creative approaches that draw on all the sources of our national power: our diversity, vibrant economy, dynamic civil society and innovative technological base, enduring democratic values, broad and deep network of partnerships and alliances.

PRESIDENT BIDEN, INTERIM NATIONAL SECURITY STRATEGIC GUIDANCE (MARCH 2021)²



² President Joseph R. Biden, Interim National Security Strategic Guidance, March 2021. <https://www.whitehouse.gov/wp-content/uploads/2021/03/NSC-1v2.pdf>



Introduction



In May 2020, the UK Government published a report³ exploring the emerging sector of Online Safety Technology described as 'Safety Tech.' A further update for 2021 was published in May 2021.⁴

Professor Mary Aiken, member of the Paladin Strategic Advisory Group was an expert advisor to the report. The report defined Safety Tech providers as those which "develop technology or solutions to facilitate safer online experiences, and protect users from harmful content, contact or conduct." (p.9)

This research identified that the UK is home to a hundred dedicated providers of Safety Tech products and services, and that this sector had grown rapidly in recent years. It highlighted some of the most innovative businesses that work globally and develop technical solutions to tackle online harms. It was also one of the first studies to define and measure the emerging value and contribution that Safety Tech makes for both economy and society.

Paladin Capital Group subsequently sought to identify comparable Safety Tech businesses and organizations active within the United States, to gauge the size, scale and demand for similar products and services nationally. The study team includes Professor Mary Aiken (Co-Lead), Sam Donaldson (Study Co-Lead, Perspective Economics), Christopher Steed, John Vernon, J.D., and Claire Haas (Paladin Capital Group).

SCOPE

Using the Safety Tech definition and building on the UK taxonomy, we have short-listed 256 dedicated and diversified providers of Safety Tech products and services active in the United States. For purposes of market analysis, we focus upon the 161 dedicated providers. However, we recognize that there are likely several providers within the US marketplace that have the relevant skills and expertise to become more involved with Safety Tech in future years.

³ UK Government (Department for Digital, Culture, Media and Sport) (2020) 'Safer Technology, Safer Users: The UK as a world leader in Safety Tech' <https://www.gov.uk/government/publications/safer-technology-safer-users-the-uk-as-a-world-leader-in-safety-tech>

⁴ <https://www.gov.uk/government/publications/safer-technology-safer-users-the-uk-as-a-world-leader-in-safety-tech>

Defining Safety Tech

This research has identified providers of Safety Tech products or services, that are active and undertake commercial activity and have established a clear presence in the United States market.

For research purposes, the following is considered within this report; however, we recognize the broader contribution of many organizations involved within the wider online safety ecosystem.

This includes organizations that:



Often work closely with law enforcement to help **trace, locate and facilitate the removal of illegal content online**;



Work with social media, gaming, and content providers to **identify criminal, harmful or toxic behavior** on their platforms;



Monitor, detect and share online harm threats with industry and law enforcement in real-time;



Develop trusted online platforms that are **age-appropriate** and provide parental reassurance for when children are online;



Verify and assure the age of users;



Actively identify and **respond to instances of online harm, bullying, harassment and abuse**;



Help organizations to **filter, block and flag harmful or illegal content** at a network or device level;



Detect and disrupt false, misleading or harmful narratives (mis- and disinformation); and



Advise and support a community of moderators to **identify and remove harmful content**.

Additionally, within the US context, we have also identified commercial organizations with technical solutions that help with:

- Identifying physical threats in real-time using technologies such as computer vision techniques, algorithms and applications for example, detecting potential shooting or violent incidents, and providing real-time responses such as notifying law enforcement or preventing access to physical sites;
- Responding to physical / public events, and using technology to enhance public safety, for example providing emergency services with event updates using crowd-sourced material; and
- Securing and protecting election integrity.

Further, the United States has a particularly active community of non-profits dedicated to tackling issues relating to online safety. Whilst this report focuses upon investment and growth of the private sector, we do note the significant contribution of these organizations.

We also note that this report reflects an initial scoping of businesses within the United States that appear to offer Safety Tech solutions, and therefore may not capture everything.

We expect that there are several hundreds of firms operating in the United States market where their product or service could have a use case applicable to making users safer. We hope that this document enables a discussion about how technology can facilitate not just safer systems, but safer platforms for users and interactions. These could include companies currently using AI, or designing solutions focused on threat intelligence or compliance that could be deployed within the context of keeping users safer online.



Safety Tech recognizes the pressing need to improve the resilience of the most critical resource in cyberspace – the human.

More importantly, technology and practice have already shown high potential to re-organize cyberspace around the needs of the human, restoring safety and control that improves productivity, while reducing the unwanted contact, conduct and content for individuals, along with the number of incidents and harms within a business.

The result is a high potential meld of technology, investment and practice that will increase the beneficial aspects and human-centered productivity of cyberspace.

CHRIS INGLIS, U.S. NATIONAL CYBER DIRECTOR AND FORMER MANAGING DIRECTOR AT PALADIN



Methodology

We set out an overview of our methodology below (and within Appendix B). This methodology is consistent with the ‘Safer Technology, Safer Users’⁵ research to enable comparisons between the relative strengths and opportunities for the sector.

METHODOLOGY	DESCRIPTION
Desk Research	Using a Grounded Theory ⁶ approach, the team reviewed over eighty pieces of academic literature, sector overviews, and grey literature. We have also longlisted more than 500 potential Safety Tech businesses within the United States and internationally to identify the characteristics and offer of Safety Tech organizations prior to shortlisting to determine a final list of Safety Tech firms operating in the United States.
Definition and Market Scoping	The research team developed a working definition for what constitutes ‘Online Safety Tech.’ This also informed the development of a ‘Safety Tech’ taxonomy within this research, and the shortlisting of Safety Tech firms, i.e., those that provide a product or service aligned to the categories.
Market Analysis	We identified relevant companies and trading information for the dedicated Safety Tech providers using Crunchbase, Pitchbook, and company websites. Investment data was sourced from Pitchbook. Please note that this analysis covers up to the end of 2020.

⁵ <https://www.gov.uk/government/publications/safer-technology-safer-users-the-uk-as-a-world-leader-in-safety-tech/safer-technology-safer-users-the-uk-as-a-world-leader-in-safety-tech>

⁶ Glaser, B., Strauss, A. (1967). The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine.



Why does Safety Tech matter?

There has never been greater focus and attention on how platforms and users can be made safer and more secure online. Further, the impact of the pandemic and the last US election continue to highlight the potential for the Internet to serve as a vehicle for disinformation, misinformation, extremist ideologies, cyber-harassment, fake news and other online harms.

Whilst a range of online platforms have implemented controls and approaches to improve user safety such as content moderation and age-verification, there is a growing recognition that a more focused and regulatory approach may be required to ensure all platforms are compliant with safer by design principles.⁷

Indeed, this is reflected in the efforts of several countries adopting new legislative and regulatory approaches to tackling online harms.

- **European Union:** The EU's Digital Service Act (DSA)⁸ regulates the obligations of digital services that act as intermediaries in their role of connecting consumers with goods, services, and content. The Act aims to protect consumers and fundamental rights online by establishing

a powerful transparency and accountability framework for online platforms that leads to fairer and more open digital markets. The Act will support compliance across the EU, support innovation across countries and ensure obligations are met that: counter illegal services, increase traceability and transparency, safeguard users, increases access to data for researchers, and support oversight structures in large and complex online spaces. It will do this by ensuring mandatory procedures are in place to increase traceability of goods, reduce vulnerability, and increase the removal of harmful material from sites.

- **Germany:** Germany's Network Enforcement Act (2017)⁹ obliges social networks to delete illegal content within a given time frame. The law dictates that where information is "manifestly" unlawful, it must be removed in 24 hours. Unlawful content is defined in the German Criminal Code and includes that which incites hatred, insult or intentional defamation. As part of the act, firms are required to complete half-yearly reporting, and where social network providers are based outside of Germany, they must have authorized staff in place to support this process. Where firms do not comply with the act, a fine up to €50m (approx. \$60m) can be administered. Notably, the German Safety Tech sector is gaining momentum.¹⁰

⁷ Australian Government, eSafety Commissioner: 'Safety by Design' Available at: <https://www.esafety.gov.au/about-us/safety-by-design>

⁸ European Commission: 'Digital Services Act' Available at: https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348

⁹ Taylor Wessing (2018) 'Germany's Network Enforcement Act and its impact on social networks' Available at: <https://www.taylorwessing.com/download/article-germany-nfa-impact-social.html>

¹⁰ Safety Tech Innovation Network (2021) 'German safety tech industry gains momentum' Available at: <https://www.safetynetwork.org.uk/articles/german-safety-tech-industry-gains-momentum>

- **Australia** introduced the world's first regulator to online safety – the eSafety Commissioner. The Australian Government has introduced the Online Safety Bill into Parliament (2021).¹¹ It covers a set of safety expectations regarding cyber abuse and harassment and take-down requirements for image-based abuse. A reformed online content scheme requiring the Australian technology industry to be proactive in addressing access to harmful online content is incorporated in the proposed Online Safety Bill. This would also expand the existing eSafety Commissioner's powers to address illegal and harmful content on websites hosted overseas, along with the capacity to disrupt access to harmful online content and to respond to crisis events (such as the viral dissemination of the Christchurch terrorist attack content online).¹²
- **United Kingdom:** The UK Government's Online Harms legislation¹³ will apply to companies that host user-generated content, search engines, and social media platforms and require these companies to tackle and remove illegal content, and to protect certain users such as children from harmful forms of content e.g., pornographic or violent content. Ofcom, the UK's communications regulator, has been appointed as the new online harms regulator. The proposed legislation imposes a duty of care on digital service providers and can fine companies that do not comply, with fines of up to £18m (approx. \$25m) or 10% of their global annual turnover (whichever is highest). The regulator will also get the power to block

access to sites. A recent report commissioned by Ofcom¹⁴ and co-led by Professor Aiken considered the role of online safety technology in tackling online harms.

The UK Minister for Digital and Culture highlighted the vital role played by the Safety Tech sector when recently announcing new laws to make the UK a safer place to be online. Under these new laws, "social media sites, websites, apps and other services which host user-generated content or allow people to talk to others online will need to remove and limit the spread of illegal content such as child sexual abuse, terrorist material and suicide content."¹⁵

Social media sites, websites, apps and other services which host user-generated content or allow people to talk to others online will need to remove and limit the spread of illegal content such as child sexual abuse, terrorist material and suicide content.

THE UK MINISTER FOR DIGITAL AND CULTURE

¹¹ Australian Government (2021) 'New Online Safety Bill Introduced' Available at: <https://www.communications.gov.au/departmental-news/new-online-safety-bill-introduced>

¹² Combating Terrorism Center (2019) 'The Christchurch Attacks: Livestream Terror in the Viral Video Age' Available at: <https://ctc.usma.edu/christchurch-attacks-livestream-terror-viral-video-age/>

¹³ TechCrunch (2021) 'UK publishes draft Online Safety Bill' Available at: <https://techcrunch.com/2021/05/12/uk-publishes-draft-online-safety-bill/>

¹⁴ Davidson, J., Aiken, M. et al (2021) 'Research on Protection of Minors: A Literature Review and Interconnected Frameworks. Implications for VSP Regulation and Beyond' Available at: https://www.ofcom.org.uk/_data/assets/pdf_file/0023/216491/uel-report-protection-of-minors.pdf

¹⁵ UK Government (2020) 'UK leads the way in a 'new age of accountability' for social media' Available at: <https://www.gov.uk/government/news/uk-leads-the-way-in-a-new-age-of-accountability-for-social-media>

The UK government is supporting the growth of the Safety Tech sector by means of the Safety Tech Innovation Network,¹⁶ a forum for Safety Tech providers to collaborate and promote their work to potential users and investors; investment in research to prototype how better use of data around online harms can lead to improved AI systems, and deliver better outcomes for citizens; organizing trade missions to priority Safety Tech export markets and collaborating across sectors to identify opportunities for innovation; support for the new Online Safety Tech Industry Association¹⁷ and the creation of a Safety Tech Providers Directory.¹⁸

Safety Tech has also started to scale up on the international stage. A Safety Tech Open Forum convened as part of the United Nations Internet Governance Forum (IGF) in November 2020, brought together a panel of experts, including Paladin Strategic Advisory Group member Professor Aiken, to debate the international potential of the sector.¹⁹ In May 2021 a Government funded initiative was launched aiming to transform data access and support Safety Tech companies to build tools to identify and remove harmful content online.²⁰

SAFETY TECH 2021



The UK Safety Tech Sector: 2021 Analysis report²¹ has found evidence of soaring sales, in what has been described as “a booming UK safety tech sector.”²²

The number of companies providing safety tech products and services has now reached 100 firms (an increase of 43%); total revenues have reached £314m in the most recent financial year; the sector’s revenue has grown by 39% in the last year; employment has grown by 29%, and additionally, there has been sustained interest in Safety Tech by the investment community with £39m raised in 2020 across 22 deals.

¹⁶ Safety Tech Innovation Network, <https://www.safetytechnetwork.org.uk/>

¹⁷ Online Safety Tech Industry Association, <https://ostia.org.uk/>

¹⁸ The Safety Tech Providers Directory, Available at: <https://www.gov.uk/government/publications/directory-of-uk-safety-tech-providers>

¹⁹ Safety Tech Innovation Network (2020) ‘Safety tech at the UN’ Available at: <https://www.safetytechnetwork.org.uk/articles/safety-tech-at-the-un>

²⁰ PUBLIC (2021) ‘Online Safety Data Initiative launches to transform data access for online harms’ Available at: <https://www.public.io/insight/insight-news/online-safety-data-initiative-launches-to-transform-data-access-for-online-harms/>

²¹ DCMS (2021) ‘The UK Safety Tech Sector: 2021 Analysis’ Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/989753/UK_Safety_Tech_Analysis_2021_-_Final_-_190521.pdf

²² DCMS (2021) ‘Sales soar for firms in booming UK safety tech sector’ Available at: <https://www.wired-gov.net/wg/news.nsf/articles/Sales+soar+for+firms+in+booming+UK+safety+tech+sector+01062021103800?open>

POSITION PAPER:

The Cyber Blue Line

A recent position paper ‘The Cyber Blue Line’²³ published by Europol, the European Union’s law enforcement agency describes cyberspace as a “new law enforcement frontier” and asks the question; “from the thin red line, to the thin blue line, to the Cyber Blue Line: Where does responsibility now lie when it comes to maintaining secure and safe societies in cyberspace?”

The paper highlights that “the economic costs of cybercrime are high, but the social costs are even higher; people are being subjected to cyber scams, to fraud and blackmail; they are being coerced, trafficked, harassed and stalked. The most vulnerable members of society, children and young people, are particularly at risk.”

Undoubtedly, the pandemic crisis has increased human psychological vulnerability and has created yet another attack vector. The paper demonstrates that “for decades, the emphasis has been on cybersecurity solutions to online threats. However, cybersecurity primarily focuses on protecting data, processes, networks and systems. It does not focus on protecting what it is to be human, what it is to be a society, and this has perhaps contributed to a protection

governance gap” pointing out that technological innovations in the new Safety Tech sector may help to address a range of online harms and crimes, from harassment to child sexual exploitation and terrorist content online. Undoubtedly, online safety technologies will be a productive future resource for law enforcement, particularly when online harms and cybercrime now have ‘Big Data’ type properties; namely, volume, velocity, and variety.

Europol, as the collective voice of European law enforcement, aims to provide a platform for important discussions by facilitating global multi-stakeholder dialogue to discuss responsibility, accountability, safety and security now mediated by technology – noting, we are all in cyberspace together.

²³ Europol (2021) ‘The Cyber Blue Line – the new law enforcement frontier’ Available at: <https://www.europol.europa.eu/newsroom/news/cyber-blue-line-%E2%80%93-new-law-enforcement-frontier>

United States: Safety Tech Policy

In the United States, there has been extensive legislative and executive activity around the degree to which online service providers can be held liable for harms occurring on their platforms and many hearings on the potential for the Internet to serve as a vehicle for disinformation, misinformation and extremist ideologies.

For example, Section 230 of the Communications Decency Act (CDA) protects online service providers from liability for hosting, removing or limiting access to user-generated content has become a significant discussion point among policy-makers and tech platforms.

“Section 230 shields certain online service providers from liability for hosting, removing or limiting access to user-generated content through two provisions. Under Section 230(c)(1), an “interactive computer service” provider, a term that includes internet service providers, online platforms and other access software providers will not be treated as a “publisher or speaker” of information content provided by a third party. Section 230(c)(2) then prevents interactive computer service providers from being held liable when they act “in good faith” to restrict or remove content the provider considers objectionable. In

other words, service providers cannot become subject to publisher liability for exercising limited editorial control over the content hosted on their sites.²⁴ Notably, a recent ruling in a US Supreme court may erode key Section 230 legal protection for social media companies.²⁵

The 116th and 117th Congress have included several meaningful hearings investigating how the United States can best seek to tackle online harms. There are currently many pieces of proposed legislation highly relevant in the context of Safety Tech, which are currently at various stages in the hearing and review process.

Some of the most notable recent examples include:

- **The PACT Act and Section 230: The Impact of the Law That Helped Create the Internet and an Examination of Proposed Reforms for Today’s Online World** (July 28, 2020);²⁶ The hearing examined the importance of Section 230 in promoting and disseminating speech online; and the history, evolution, and expansion of Section 230’s protections for online platforms. It offered an opportunity to discuss ways to ensure platforms are accountable for their content moderation practices and what legislative measures (e.g., transparency and accountability tools) can empower consumers online.

²⁴ Arnold & Porter (2020) ‘Beyond the Headlines: Analysis of the Executive Order on Preventing Online Censorship and the Potential Impact on Section 230 Policy’ Available at: <https://www.arnoldporter.com/en/perspectives/publications/2020/06/analysis-of-the-eo-on-preventing-online-censorship>

²⁵ Forbes (2021) ‘Ruling Against Facebook In Sex Trafficking Case Threatens Key Legal Shield For Social Media Platforms’ Available at: <https://www.forbes.com/sites/graiondangor/2021/06/25/ruling-against-facebook-in-sex-trafficking-case-threatens-key-legal-shield-for-social-media-platforms/?sh=3e66f5833acc>

²⁶ U.S. Senate (2020) *The PACT Act and Section 230: The Impact of the Law that Helped Create the Internet and an Examination of Proposed Reforms for Today’s Online World*. 106 Dirksen Senate Office Building: U.S. Senate Committee on Commerce, Science, & Transportation. Available at: <https://www.congress.gov/event/116th-congress/senate-event/327722?s=1&r=3>.

- **Emerging Trends in Online Foreign Influence Operations: Social Media, COVID-19, and Election Security** (June 18, 2020);²⁷ Hearing to examine social media companies' strategies and efforts to combat election interference, public health disinformation related to the outbreak of novel coronavirus, known as COVID-19, and foreign influence operations on online platforms.
- **A Country in Crisis: How Disinformation Online Is Dividing the Nation** (June 24, 2020);²⁸ This hearing focused specifically on the impact of COVID-19 and the death of George Floyd. It noted the danger of widespread disinformation, which may be exacerbating division in the US.
- **Countering Online Harms Act** (May 19, 2020);²⁹ this act would require the Federal Trade Commission to study and report on how artificial intelligence may be used to identify, remove, and take action to address online harms content, furthering other illegal activity such as the sale of opioids, child sexual exploitation, terrorism and the sale of counterfeit products.
- **Earn It Act: Holding the Tech Industry Accountable in the Fight against Online Child Sexual Exploitation** (March 11, 2020);³⁰ This hearing was used to flag the number of reports made to the National Center for Missing and Exploited Children Cyber Tipline in 2019, which include 15.8m from Facebook, 46k from Twitter, and notably, 8 on Amazon.³¹ The disparity in the number of reports made has been attributed to the liability protection awarded under Section 230.
- **Americans at Risk: Manipulation and Deception in the Digital Age** (January 08, 2020);³² This hearing before the Subcommittee on Consumer Protection examined fraudulent, deceptive, and manipulative practices used on Internet platforms and social networks, and reviewed strategies and recommendations to combat deceptive and manipulative online practices.

²⁷ U.S. House of Representatives (2020c) *Emerging Trends in Online Foreign Influence Operations: Social Media, COVID-19, and Election Security*. Washington, D.C.: U.S. House Intelligence Committee. Available at: <https://www.congress.gov/event/116th-congress/house-event/110805> (Accessed: 18 April 2021).

²⁸ U.S. House of Representatives (2020a) *"A Country in Crisis: How Disinformation Online Is Dividing the Nation"*. Washington, D.C.: U.S. House Energy and Commerce Committee. Available at: <https://www.congress.gov/event/116th-congress/house-event/110832> (Accessed: 18 April 2021).

²⁹ Guthrie, B. (2020) *Countering Online Harms Act*. Available at: <https://www.congress.gov/bill/116th-congress/house-bill/6937> (Accessed: 18 April 2021).

³⁰ Graham, L. (2020) *Text - S.3398 - 116th Congress (2019-2020): EARN IT Act of 2020*. Available at: <https://www.congress.gov/bill/116th-congress/senate-bill/3398/text> (Accessed: 4 December 2020).

³¹ <https://www.judiciary.senate.gov/chairman-graham-statement-for-the-record-at-hearing-titled-the-earn-it-act-holding-the-tech-industry-accountable-in-the-fight-against-online-child-sexual-exploitation>

³² U.S. House of Representatives (2020b) *Americans at Risk: Manipulation and Deception in the Digital Age*. 2123 Rayburn House Office Building, Washington, D.C.: U.S. House Energy and Commerce Committee. Available at: <https://www.congress.gov/event/116th-congress/house-event/110351> (Accessed: 18 April 2021).

Most recently (Feb 7, 2021) U.S. Senators. Mark R. Warner (D-VA), Mazie Hirono (D-HI) and Amy Klobuchar (D-MN) announced the **Safeguarding Against Fraud, Exploitation, Threats, Extremism and Consumer Harms (SAFE TECH) Act**³³ aiming to “reform Section 230 and allow social media companies to be held accountable for enabling cyber-stalking, targeted harassment, and discrimination on their platforms.”³⁴

Section 230 of the CDA is often cited as the most critical law supporting the Internet, e-commerce and the online economy. However, it continues to be subject to intense criticism

regarding the current proliferation of hate speech, defamation and disinformation online.³⁵ Given the ever-increasing complexity and prevalence of the Internet in shaping political opinion and influencing society’s response to key events, most recently COVID-19, Black Lives Matter protests and the election campaign, not to mention the reported role of social media and private-messaging apps underpinning recent violent events on Capitol Hill, there is an urgent need to focus on, invest in, and support Safety Tech solutions to all forms of technology mediated harms, and indeed work towards mitigating or preventing real-world associated harms.



The future security of our nation as a whole, and its citizenry down to the person, will depend greatly on investments into technologies that counter ‘online harms’ combined with the development of curricula that raises the digital literacy of the general population. National Security and personal security are both at stake. We especially need to equip primary school teachers (K-12) with age-appropriate content that can illuminate how the technology actually works. For example, develop educational resources that would help to combat the mis/disinformation problem, by explaining how the Internet and algorithms work, and in doing so improve critical thinking skills around Internet usage to prevent online harms.

VICE ADMIRAL (RET.) JAN TIGHE FORMER VICE ADMIRAL, SERVING AS THE DEPUTY CHIEF OF NAVAL OPERATIONS FOR INFORMATION WARFARE AND AS THE 66TH DIRECTOR OF NAVAL INTELLIGENCE. MEMBER OF PALADIN STRATEGIC ADVISORY GROUP



³³ Warner, M. R. (2021) Safeguarding Against Fraud, Exploitation, Threats, Extremism, and Consumer Harms Act. Available at: <https://www.congress.gov/bill/117th-congress/senate-bill/299/titles> (Accessed: 18 April 2021).

³⁴ Warner, M. R. (2021) ‘Warner, Hirono, Klobuchar Announce the SAFE TECH Act to Reform Section 230’ <https://www.warner.senate.gov/public/index.cfm/2021/2/warner-hirono-klobuchar-announce-the-safe-tech-act-to-reform-section-230>

³⁵ National Law Review (2020) The Communication Decency Act and the DOJ’s Proposed Solution: No Easy Answers’ Available at: <https://www.natlawreview.com/article/communication-decency-act-and-doj-s-proposed-solution-no-easy-answers>

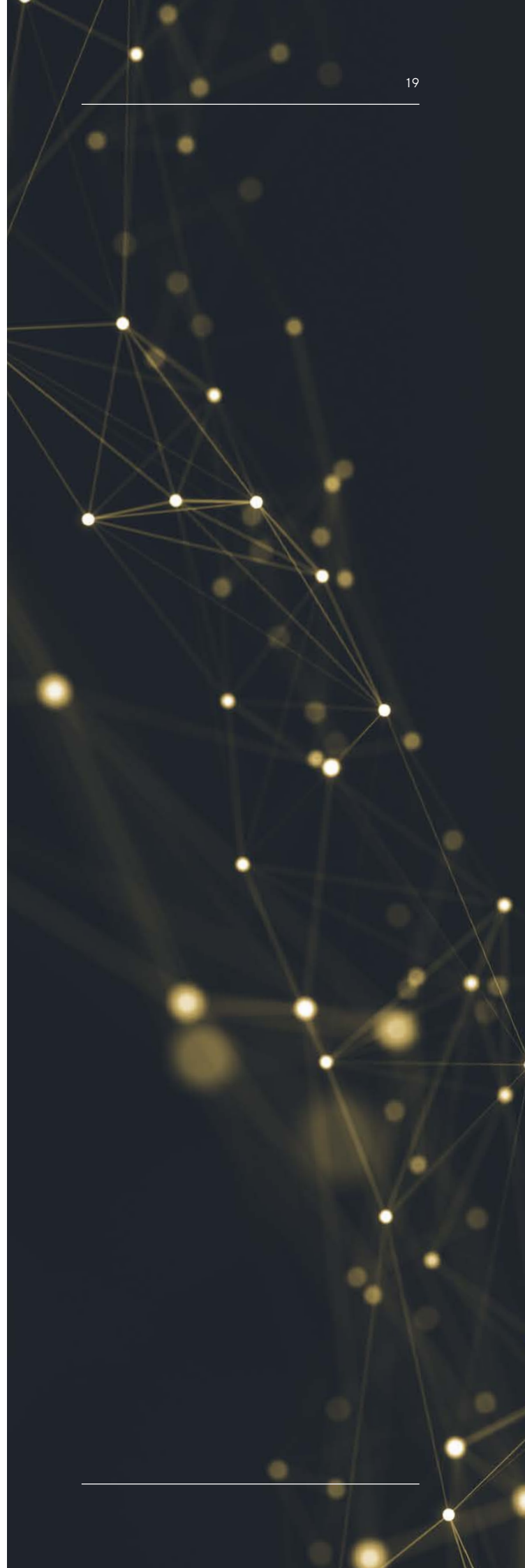
Safety Tech Taxonomy

Safety Tech is a fast-moving environment, with respect to the harms, risks and technologies in scope. Each of the organizations identified are involved in developing novel approaches to identify and counter risks before users are harmed, and many of the organizations will offer multiple products and services accordingly.

This taxonomy, therefore, reflects an overview of the current products and services offered by Safety Tech providers to inform analysis of the respective size and scale of the sector (2020-21) and to identify sector strengths and opportunities.

The research team has developed this taxonomy for such purposes. It is not considered exhaustive, and we recommend that it is revisited on an ongoing basis to reflect the pace with which this sector is developing new solutions to existing and evolving harms.

Please note that the order of the sub-categories does not imply any ranking regarding importance, size, or scale.





01 AT SYSTEM LEVEL

Automated identification and removal of illegal imagery

- Use of technology to identify and remove known illegal child sexual exploitation and abuse (CSEA) and terrorist content, especially imagery and video, frequently through the use of hash lists.



02 AT PLATFORM LEVEL

Supporting human moderators

- Identifying and flagging to human moderators for action:
 - Potentially illegal content or conduct, such as hate crimes, harassment or suicidal ideation
 - Harmful content or conduct which breaches site T&Cs, such as cyberbullying, extremism advocacy of self-harm and extreme violence content
- Reducing moderators' own exposure to harmful content.

Enabling age-appropriate online experiences

- Use of age-assurance and age-verification services to limit children's exposure to harmful content.



03 AT DEVICE OR ENDPOINT LEVEL

User-initiated protection

- User, parental or device-based products that can be installed on devices to help protect the user from harm.

Network filtering

- Products or services that actively filter content, through black-listing or blocking content perceived to be harmful. This can include solutions provided to schools, businesses or homes to filter content for users.



04 IN THE INFORMATION ENVIRONMENT

Identifying and mitigating disinformation

- Flagging of content with false, misleading and/or harmful narratives, through the provision of fact-checking and disruption of disinformation (e.g., flagging trusted sources).
- Technology to safeguard / protect election integrity.



05 AT CYBER-PHYSICAL LEVEL

Using technology to identify, prevent, and mitigate harmful incidents from occurring, or respond to events

- Use of AI / video analytics to identify events and or threats in real-time in the real-world e.g., identification of weapons in a public setting.
- Technology to aid emergency response / public safety in the physical domain.

System Level Factors: This refers to organizations involved at the highest levels, often working closely with law enforcement, to trace, locate and remove (or help to facilitate the removal) of illegal content online.

System-Wide Governance: These organizations help to identify and tackle some of the Internet's most harmful content e.g., child sexual abuse and exploitation, and terrorist content.

This can be achieved through:

- Working closely with law enforcement to assist with investigative capabilities e.g., use of digital forensics to scan for known illegal content using MD5 hashes;
- Maintaining and providing access to technology aimed at preventing the upload, facilitating the removal of illegal content; and
- Combating abuse or threats with automated content analysis and AI e.g., automated detection of terrorist content, including previously unseen material.

Platform Level Factors: This refers to organizations that are involved in making online services safer and typically work at the platform level i.e., work alongside social media, gaming, and content providers to improve safety and behavior within their platforms, for example; protecting gamers from online harms, fraud and toxicity.

These are segmented into the following sub-categories:

Platform Governance: These organizations are focused upon helping providers of online content to govern their offering with respect to illegal content. Whilst there is some overlap with 'System Governance', this is more focused upon organizations that help to tackle issues such as:

- Embedding prevention mechanisms e.g., using machine learning to prohibit the production of indecent underage imagery on social media platforms; and
- Identifying and blocking harmful images and videos in real-time;
- Identifying advanced threats; and
- Helping companies protect their employees from online harassment and abuse.

Platform Moderation & Monitoring: These organizations also help providers of online content to monitor and moderate behavior and content posted within their platforms. This is typically focused upon reducing harmful content or behavior e.g., offensive language, bullying, harassment or toxic content.

This can include:

- Moderation and monitoring of content e.g., pre-moderation or post-moderation of content, undertaken by automated content analysis and / or humans;
- Chat / gaming moderation e.g., identifying and removing users subject to language or words used; and
- Behavioral Monitoring e.g., identifying good and bad behavior, and tackling cyber-harassment or hate speech, typically using Natural Language Processing within online communities.

Age Orientated Online Safety: These organizations seek to support online content providers in ensuring that their platforms are either age-appropriate and increase the privacy of children online (e.g. compliant with COPPA,³⁶ or that the content and access requirements are suitable should the website or app be targeted at under-13s/ under-18s, thereby ensuring 'safety by design'), or provide age-assurance services (i.e. help companies to validate and confirm that only particular age groups can access specific content).

³⁶ Children's Online Privacy Protection Rule ("COPPA") <https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule>

Endpoint Level Factors: This refers to organizations that provide products or services that help to ensure that the device being utilized by the end-user is suitably secure with respect to online safety. This focuses upon online safety solutions (i.e., ensuring that the user's risks with respect to content, conduct, and contact are reduced). It does not include endpoint protection from viruses, malware, or adware – which are covered by the field of cybersecurity.

User protection at the endpoint level can be segmented into two main categories:

User-initiated protection (user, parental or device-based): This includes organizations that provide products or services that can be installed on devices to help secure the end-user from online harms (typically a parent or guardian installing on behalf of a child). The underlying ambition is to create a safer online experience for the user e.g., through safeguarding assistants, oversight of social media content, or through monitoring of a child's digital or online behavior and interaction with other users. Where deployed, these solutions can help to prevent issues relating to sexting, bullying, abuse, or aggression.

Network filtering: This includes organizations involved in providing products or services that actively filter content (e.g., through white-listing or black-listing, actively blocking content perceived as harmful or illegal). This can often include solutions provided to schools or home users to filter content for users.

Information Environment Level Factors: This refers to organizations that actively detect and disrupt false, misleading and / or harmful narratives. This includes tackling misinformation and disinformation through the provision of fact checking and disinformation research and disruption. Organizations within this space seek to ensure citizen information accuracy and facilitate trust in the information environment and wider society. Further, this includes organizations using technology to help safeguard and protect election integrity.

Cyber-Physical Level Factors: This refers to organizations developing technology to identify, prevent, and mitigate harmful incidents, or to respond to events. This includes, for example, the use of AI / video analytics to identify events and or threats in real-time in the real-world e.g., identification of weapons in a public setting, or technology to aid emergency response / public safety in the physical domain, for example, active school shooting or workplace violence events. This category is new to the US study and builds on previous taxonomies as the researchers have identified several providers of physical public safety solutions.

Other

Online Safety Professional Services: This includes organizations typically involved in supporting the design, implementation and testing of online safety by providing compliance services, research, frameworks and methodologies for auditing, evaluating or mitigating potential harms, and helping to enable the development of safer online communities.

Support: Further, this analysis has also sought to identify organizations involved in supporting the development and scaling of online safety products and services but do so in an advocacy capacity for example, charities, NGOs and thinktanks.

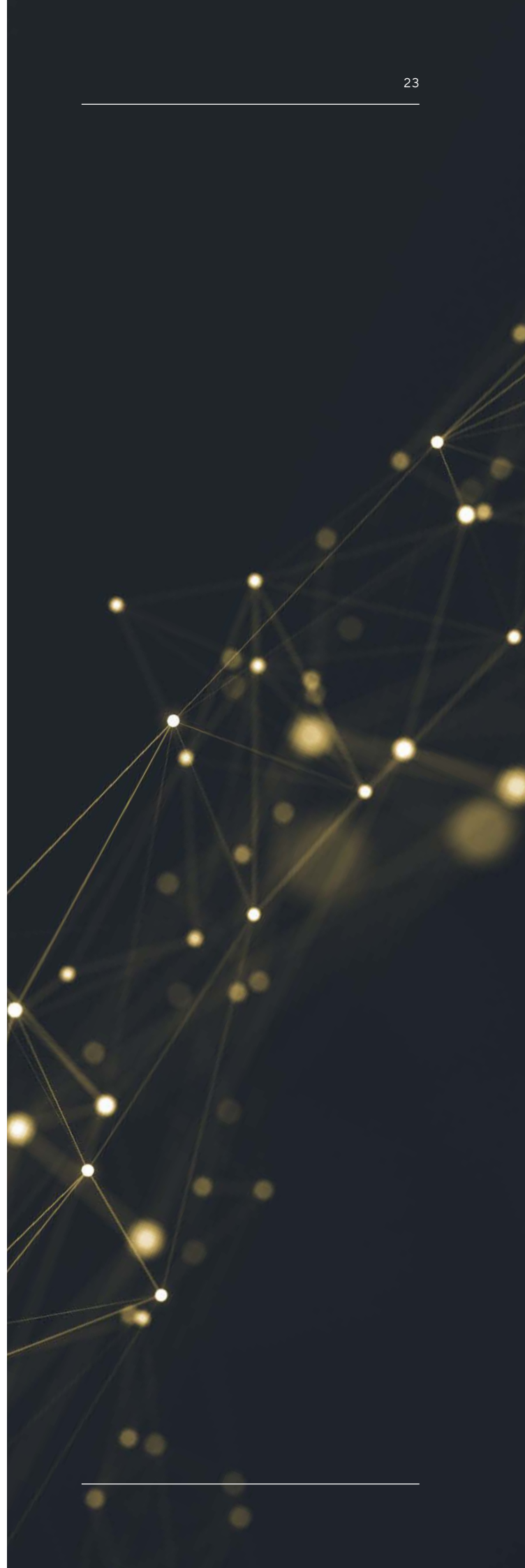
The US Safety Tech Market

This chapter sets out a summary of the market profile of the US Safety Tech sector. It identifies Safety Tech providers, which:

- ✓ have a clear presence in the US market
- ✓ demonstrate an active provision of commercial activity related to online safety technology (e.g., through the presence of a website / social media)
- ✓ provide Safety Tech products or services to the market (i.e., sell or enable the selling of these solutions to other customers)
- ✓ have identifiable revenue, employment, or trading activity within the US.

Please note that we also recognize that several of the larger tech companies are involved in the production or development of Safety Tech solutions (e.g., Microsoft's PhotoDNA, AWS' Rekognition Image Moderation API, Facebook and Google providing access to OpenAI and TensorFlow etc.). However, these providers are not measured within this exercise, which is focused on dedicated third-party Safety Tech providers.

We also note that this report reflects an initial scoping of businesses within the United States that appear to offer Safety Tech solutions, and therefore may not capture everything. Using the Safety Tech definition and taxonomy, we have identified 161 commercial organizations based in the United States dedicated to providing relevant Safety Tech products and services.



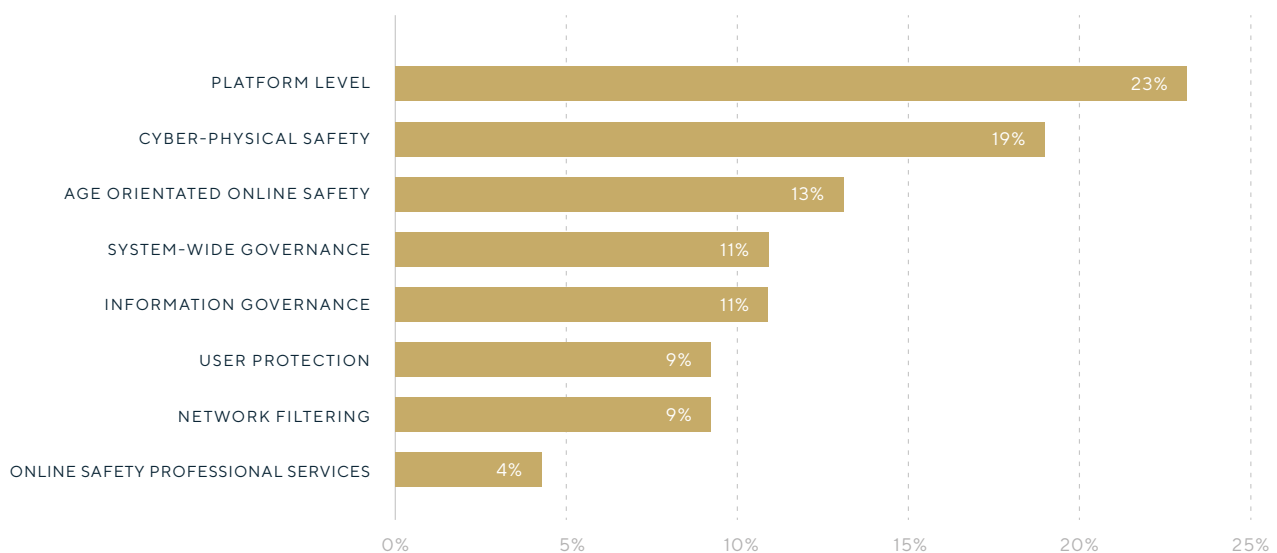
Products and Services

An abstract geometric pattern of interconnected lines and dots, resembling a network or constellation, is overlaid on a dark background. The lines are thin and light-colored, connecting several prominent white dots and many smaller, fainter dots. The pattern is more dense in the upper right and lower left areas, with lines radiating from central points.

For the dedicated organizations identified for analysis, we have identified company descriptions of what they offer using website and trading data available. The nature of this sector means that some organizations provide a wide range of products and services – for example, content moderation, sentiment analysis, and advisory services.

However, we have identified the **best fit** of each of the commercial organizations against the taxonomy classification and sub-categories, to illustrate the overall sector composition.

PRODUCT & SERVICE CLASSIFICATION:



Source: Perspective Economics (n = 161)

Overall, this suggests that the United States is home to a wide range of product and service capabilities within the Safety Tech marketplace. As noted within a similar study exploring the UK market, this is important as it demonstrates a wide-reaching pool of expertise and capability to protect and mitigate against a range of online harms.

However, we note that distinctively, the United States market appears to place substantial focus upon 'platform level' protection i.e., there are

several content moderation and anti-toxicity / behavioral solutions offered on a Business to Business (B2B) / Software-as-a-Service (SaaS) basis; and somewhat differentiated from the UK market in terms of providing a range of technology based solutions regarding physical safety, such as video analytics of targeted violence or terrorism (TVT)³⁷ threats within crowds or schools, or enabling improved emergency response to physical harms using online intelligence.

³⁷ Department of Homeland Security (2019) 'Strategic Framework for Countering Terrorism and Targeted Violence' Available at: https://www.dhs.gov/sites/default/files/publications/19_0920_plcy_strategic-framework-countering-terrorism-targeted-violence.pdf

EXAMPLE SAFETY TECH ORGANIZATIONS ACTIVE IN THE UNITED STATES

System-Wide Governance

- **AccessData:** Founded in 1987, AccessData provides digital forensics software solutions for law enforcement and government agencies, including the Forensic Toolkit (FTK) Product.
- **Magnet Forensics** builds software that automates the recovery of digital forensic evidence.
- **Hubstream, Inc.** provides investigative intelligence and case management solutions. They are a partner of Project VIC (a coalition of public sector, NGO and private sector organizations) working together to help law enforcement agencies to rescue child victims of abuse and stop offenders.
- **DarkOwl** extracts and analyzes data at scale from the Darknet, helping to tackle the industry of 'Disinformation-as-a-Service' on the Darknet.
- **Nisos, Inc** provides managed intelligence solutions for organizations to secure from cyber threats, and detect and respond to disinformation, advanced threats, and platform abuse.
- **GamerSafer** helps multiplayer games scale online safety, positive and fair play experiences to millions of players. It enables platforms to prevent online harm, fraud and toxicity using computer vision and artificial intelligence solutions.
- **OpenSlate** provides independent, industry-standard metrics to foster a healthy online video ecosystem and support high-quality content. Its content rating system ingests and analyzes more than 500 million videos every day, providing advertisers with a third-party assessment of the content that they support with their ad dollars.
- **ModSquad:** With operation centers in Northern California, Austin and Northern Ireland and offices in San Francisco, London and New York, ModSquad has over 10,000+ mods globally helping companies to moderate content, chat with customers, and manage communities.
- **Samurai Labs:** With offices in Poland and the US, Samurai helps keep children and online communities safe by detecting and responding to cyberviolence, sexual predators, suicidality and crisis events.
- **CleanSpeak** provides intelligent and scalable profanity filtering and moderation software.

Platform Level

- **Confiant's** ad security & ad quality solution protects publishers and ad platforms from security, quality, and suitability threats.
- **Spectrum Labs'** content moderation, contextual AI system, and data labelling help combat online toxicity.
- **Tall Poppy** builds tools and services to help companies protect their employees from online harassment and abuse.
- **SuperAwesome:** Recently acquired by Epic Games, SuperAwesome's kidtech is used to enable safe engagement with the global under-13/under-16 audience. Used by hundreds of brands and content owners, SuperAwesome's technology provides the tools for safe digital engagement with almost half a billion kids every month.
- **TotallyKidz** provide creative solutions across a brand safe, COPPA compliant gaming network since 2011. They help brands deliver results, reach and engage kids as they seek out cross-screen entertainment globally.

- **Gabb Wireless** provides kids with their 'perfect first phone.' It has no Internet browser, app store, or social media – and is intended to be completely safe for kids with no parental controls needed.

Cyber-Physical Safety

- **Vector LiveSafe** is a risk intelligence platform that surfaces early warning insights and prevents serious safety and security incidents to mitigate operational risks, reduce financial losses, and make places safer for people to work, learn, and live.
- **Jiobit** develops advanced location monitoring technology for families, to help share child locations with trusted adults.
- **bSafe** has developed technology for mobile phones and alarm centers to prevent crimes such as violence and sexual assault, as well as gathering evidence in cases where it has occurred.
- **Athena Security** develops solutions for temperature detection to help with finding fevers in search of COVID-19, as well as gun detection to protect employees or students before an active shooter situation is initiated.

User Protection

- **Circle** provides parental controls, time limits and location tracking through mobile devices.
- **Qustodio** helps monitor and track online activity, block dangerous sites and protect kids from online bullying.
- **WebSafety** assists parents in monitoring their children's mobile device usage and activities.
- **NetSanity** gives parents control over the access their kids have to their mobile devices.
- **NetNanny** provides parental control and web filtering software, to provide visibility and control over kids' online activity.

Network Filtering

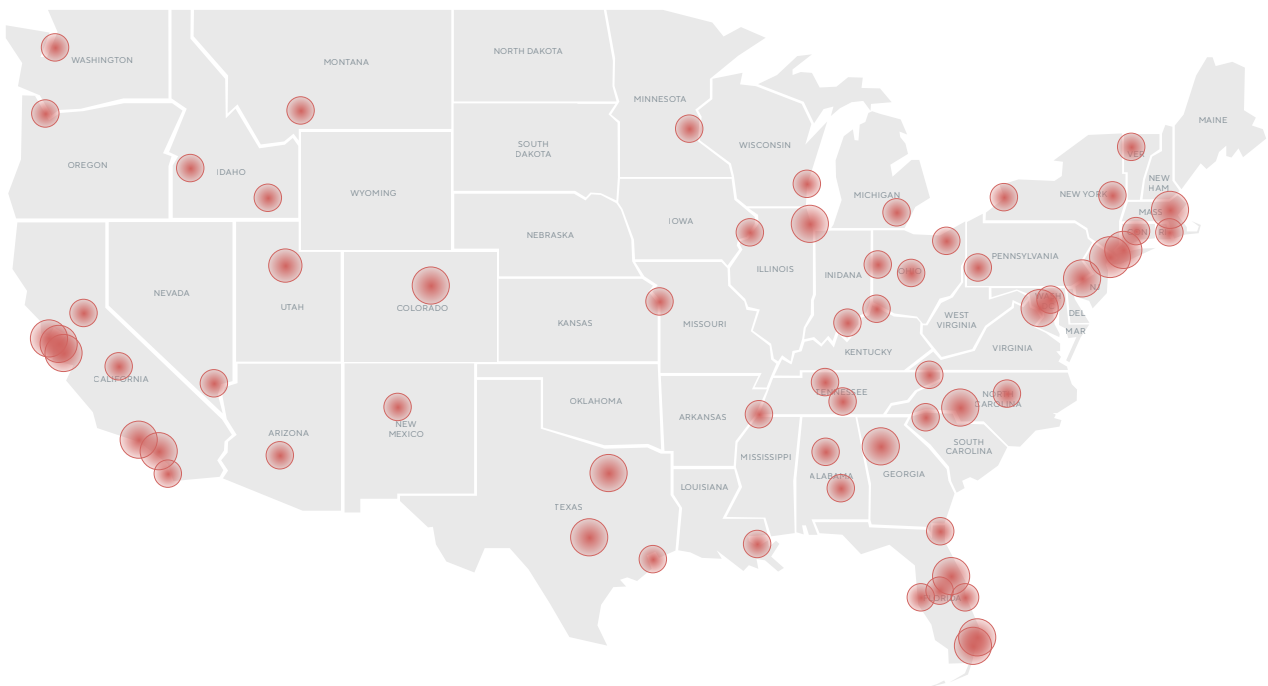
- **Zvelo** provides comprehensive URL categorization and malicious detection solutions for web content, traffic, and connected devices.
- **Securly** helps secure schools and homes, through cloud-based web-filtering, AI-based scanning of content, and visibility of online activity such as notifications of threats, cyberbullying, suicide, and nudity.
- **DNSFilter** provides content filtering and threat protection, to help stop users from viewing inappropriate or undesirable content.
- **SafeDNS** provides cloud-based web filtering to help safeguard families and organizations against threats and objectionable content on the Internet.
- **CIPAFILTER's** real-time context-sensitive filtering across games, proxies, and sexual content helps to keep students safe online.
- **NetSpark's** content filtering tool focuses on enabling access to content, not just blocking it. It contextually understands web content, prior to filtering out any inappropriate elements.

Information Governance

- **Truepic** develops secure camera technology for mobile devices. Truepic Vision enables partners to instantly gather and view trusted visual documentation from anywhere in the world. It is dedicated to restoring trust in every pixel of consequence, with the goal of having a shared sense of visual reality across the Internet by 2030.
- **NewsGuard** labels news sources with an indication of general trustworthiness and whether it has a history of running stories labelled as fake news.
- **Blackbird.AI** helps brands, governments and citizens detect and understand disinformation.

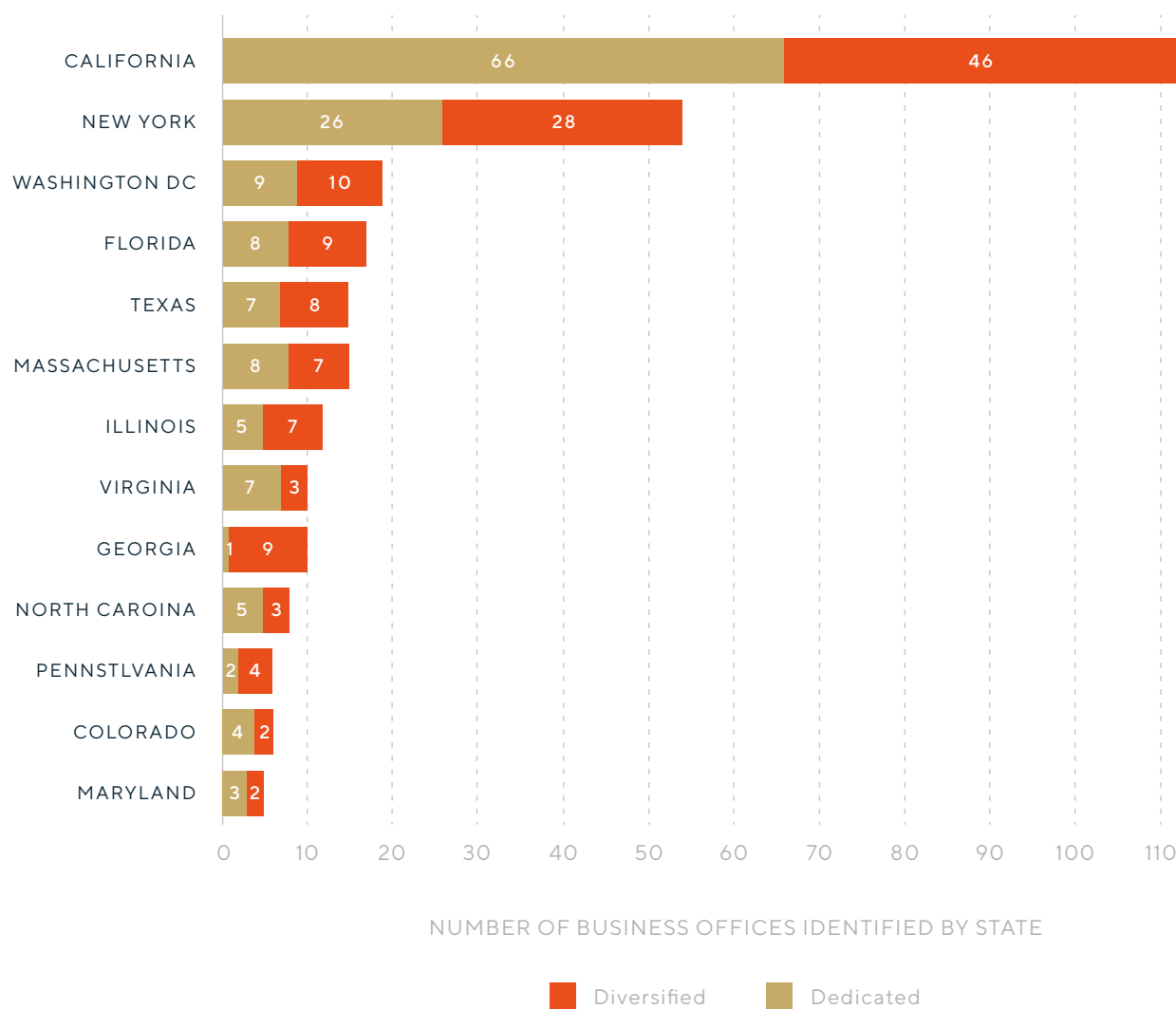
Location

The map below sets out the 337 office locations identified, covering the 256 dedicated **(161)** and diversified **(95)** providers of Safety Tech products and services active in the United States.



Source: Perspective Economics, Google Maps API (N = 337 offices)

The chart below sets out the count of offices at the state level (those with more than five offices), demonstrating that a third of firms have a presence in California, followed by New York, Washington DC, Florida, and Texas.



Source: Perspective Economics

Number of Safety Tech Companies: Time Series

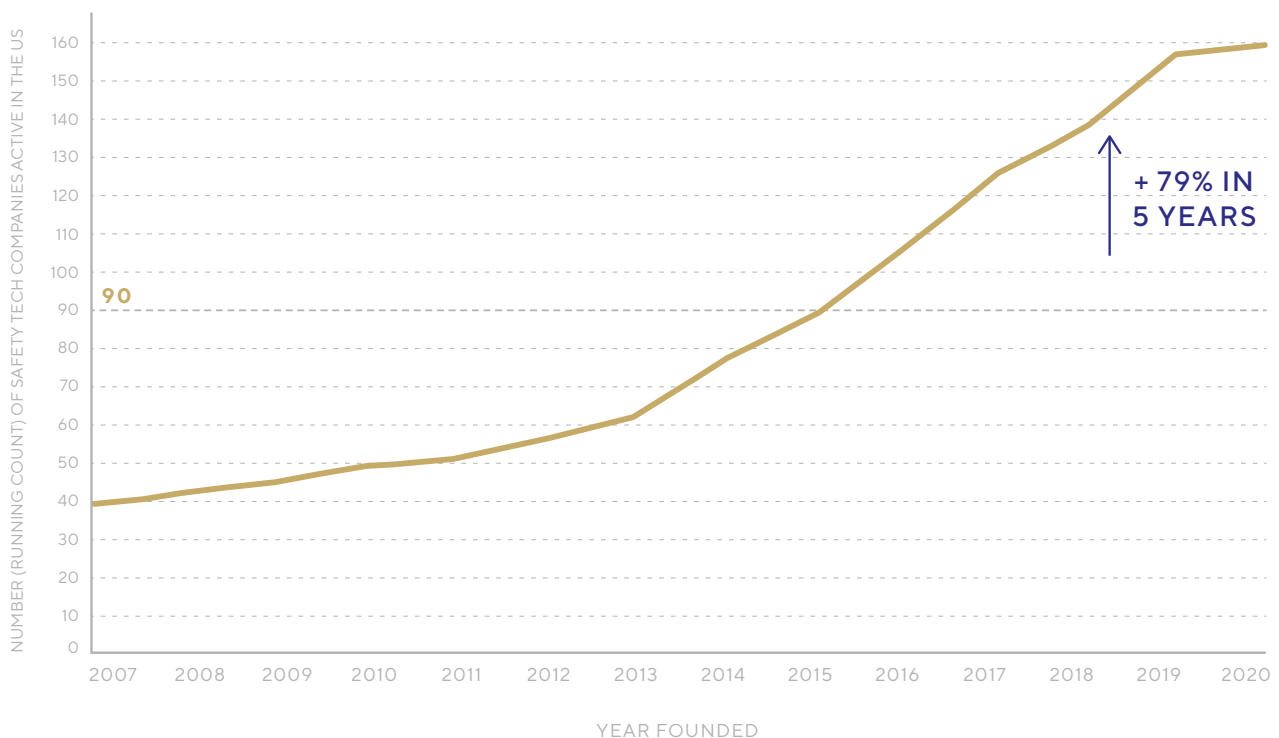
We have identified the founding year of each of the dedicated Safety Tech businesses within the analysis by time series.

The mid-1990s witnessed the emergence of parental control software and Internet filtering. However, since the mid-2000s, new providers have emerged in the marketplace with a greater focus on areas such as content moderation, age-appropriate online experiences, digital forensics and identification of illegal and harmful material.

The graph below sets out how, since the end of 2014, the number of dedicated Safety Tech firms has rapidly increased from 90 to 161 by the start of 2020 (an increase of 79% in five years). In other words, there has been a considerable increase in the number of new companies being formed to address online harms.

Of these new businesses formed within the last five years, these typically offer products or services aligned to the taxonomy as per graphic on right.

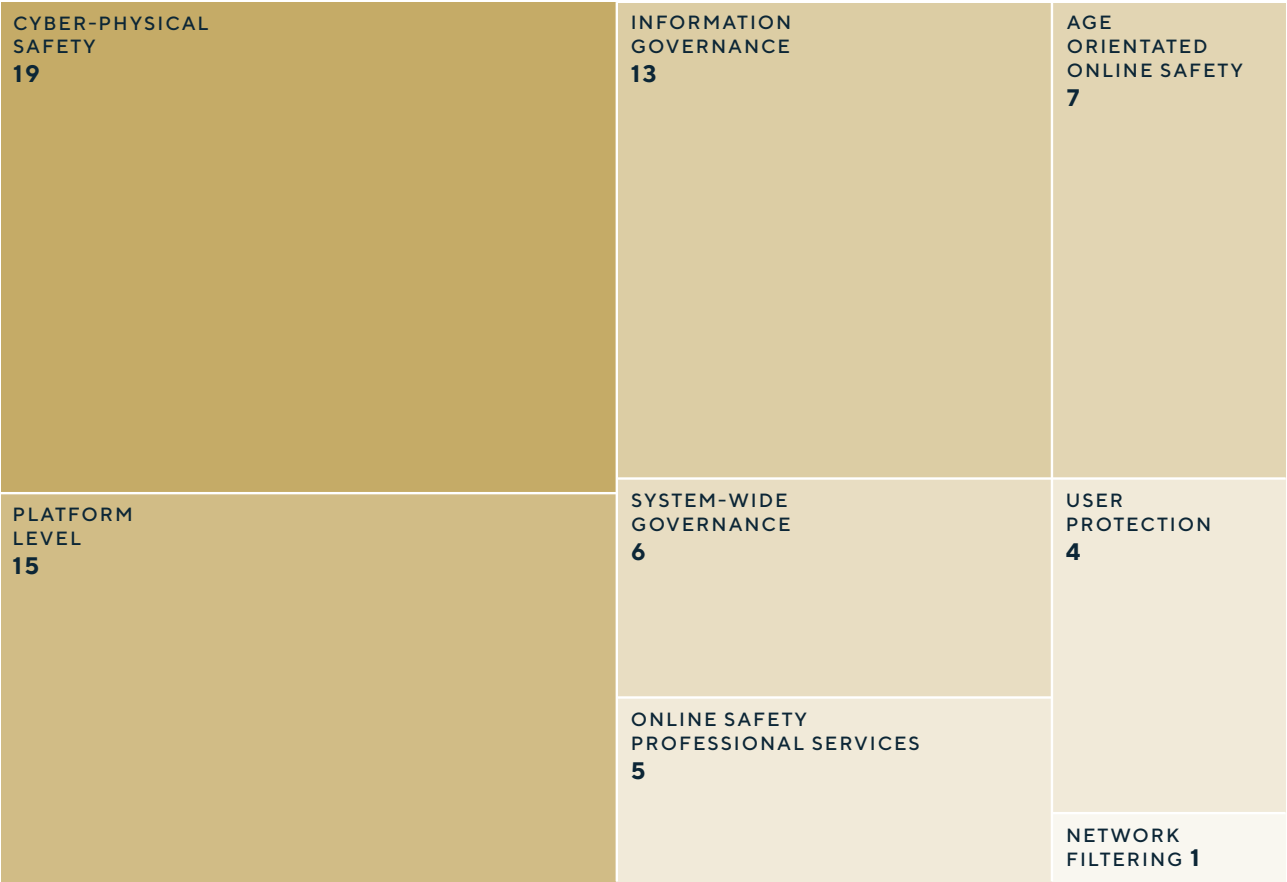
NUMBER OF ACTIVE SAFETY TECH COMPANIES BY YEAR FOUNDED



This demonstrates that there has been considerable growth in the number of new start-ups using AI for physical safety, tackling

disinformation and ‘fake news’ (particularly since entering the lexicon in 2016), platform level content-moderation and anti-toxicity solutions.

NUMBER OF COMPANIES FOUNDED SINCE 2016 – BY MAIN PRODUCT OR SERVICE OFFERING:



Source: Perspective Economics



Safety Tech is an idea whose time is coming on both sides of the Atlantic. If we succeed in making the technology itself safer, cyber security becomes so much easier and we can focus on the really hard stuff.

CIARAN MARTIN, MANAGING DIRECTOR AT PALADIN CAPITAL GROUP, PROFESSOR OF PRACTICE IN PUBLIC MANAGEMENT AT THE UNIVERSITY OF OXFORD'S BLAVATNIK SCHOOL OF GOVERNMENT, AND FORMER CHIEF EXECUTIVE OF THE UK'S NATIONAL CYBER SECURITY CENTRE



Estimated Employment

For the 161 dedicated Safety Tech firms, we have reviewed company-level data to estimate the current number of full-time equivalent staff working in US Safety Tech.

This includes staff working for these companies within the United States but does not include outsourced staff or those working outside of the United States.

We estimate that there are currently 8,800 professionals working in Safety Tech in the United States. This is over four times larger than the workforce identified within the United Kingdom.

Some of the largest employers include ModSquad (estimated 600 staff in the US, with over 10,000 globally), Magnet Forensics (>250 staff), GoGuardian (>200 staff) and Securly (>150 staff).

Investment in US Safety Tech

INTRODUCTION

This section sets out an overview of the investment landscape for the firms identified within this analysis. All of the organizations identified were matched with Pitchbook and Crunchbase Pro to identify announced investments to inform this analysis.

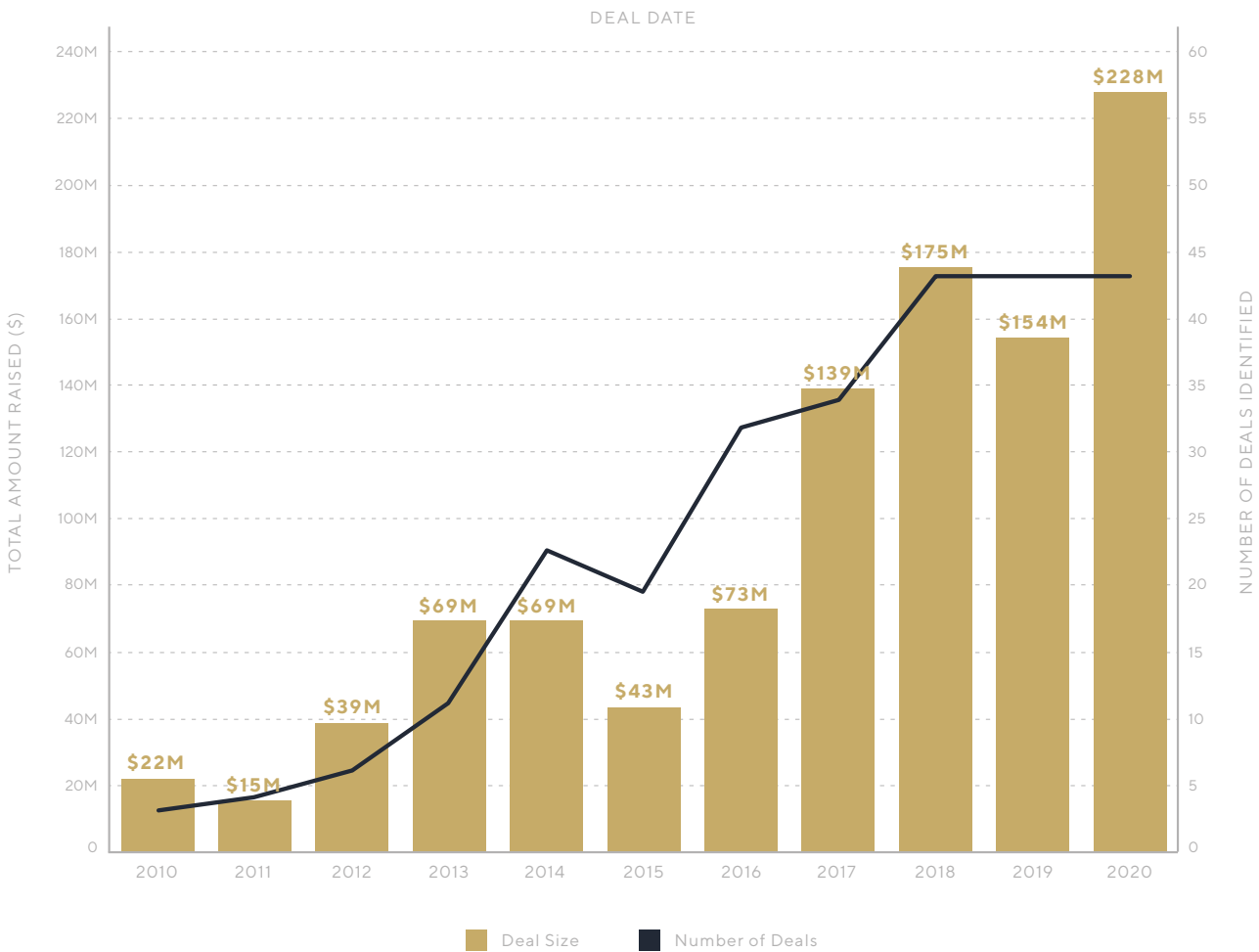
IDENTIFIED INVESTMENTS TO DATE

Of the dedicated Safety Tech organizations identified, 85 (53%) have raised some form of external investment. **In total, these businesses have raised over \$1bn in external investment since 2010.**

The chart breaks down the value and volume of investments made in dedicated Safety Tech firms in the United States since 2010. This demonstrates a significant uptick in investment activity since the start of 2017, with nearly \$700m raised in the last four years alone.

Further, despite challenging market conditions, 2020 was a record year for Safety Tech investment in the US, with \$228m identified in announced VC investments.

INVESTMENT BY YEAR (VALUE AND VOLUME) IN DEDICATED SAFETY TECH COMPANIES

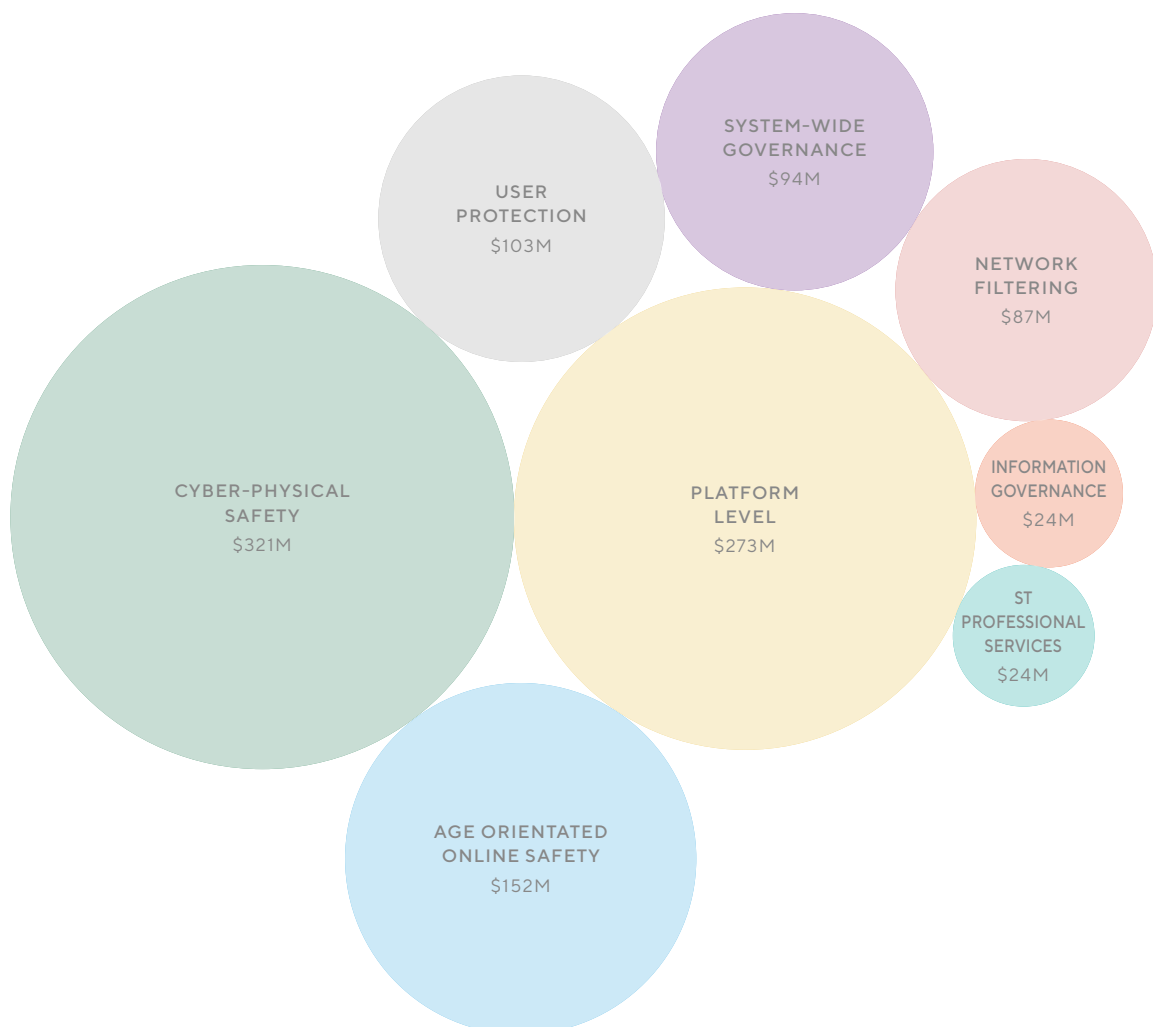


Source: Pitchbook, Perspective Economics

Additionally, we have segmented total investment by each of the Safety Tech product and service sub-categories below. This demonstrates the significant importance of technical solutions within the categories of supporting physical safety, improving content moderation, monitoring within online platforms, and supporting platforms with age assurance and regulatory compliance.

There are also emerging signals of increased investment attention in firms focused on tackling online harms such as disinformation, improving brand safety, and ID and age verification.

INVESTMENT BY TAXONOMY CATEGORY IN DEDICATED SAFETY TECH COMPANIES:

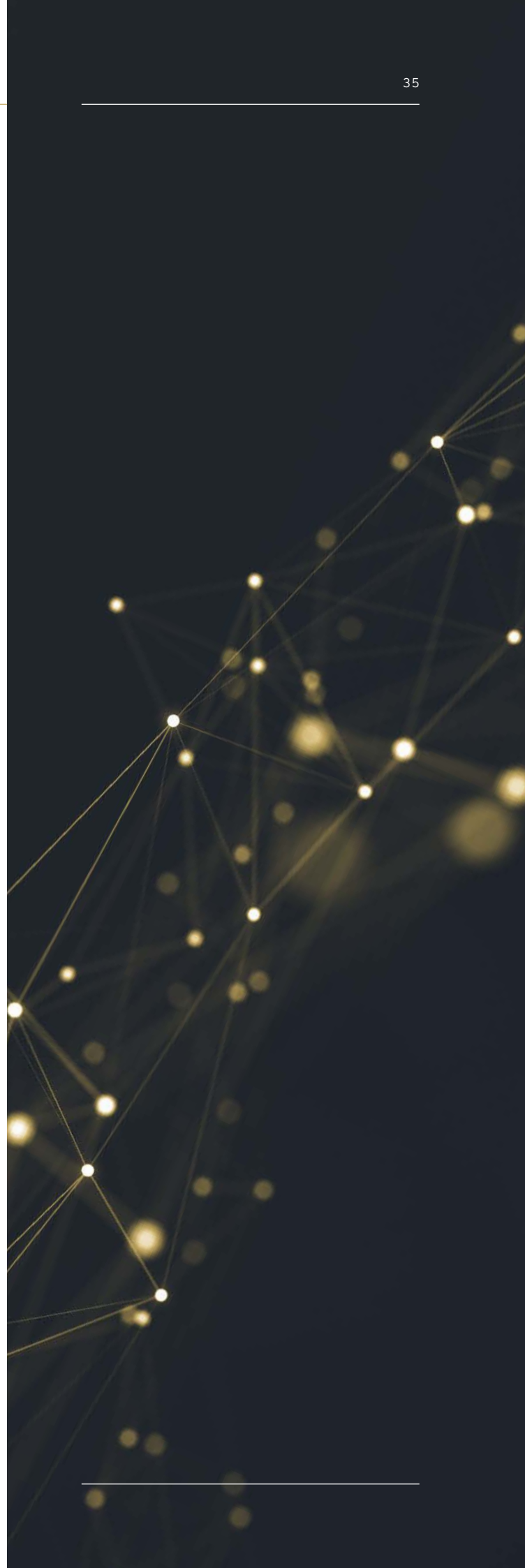


Source, PE, Crunchbase Pro





It is imperative that as we become ever more dependent on the Internet we find new ways of reducing the online harms that threaten our businesses, our families and the vulnerable in society. This timely and ground-breaking Paladin report highlights the US investment opportunity offered by the rapidly growing new field of 'Safety Tech' that is starting to provide innovative solutions to combat malign online activity, not least the digital subversion, abuse and harassment too often corrupting social media.

PROFESSOR SIR DAVID OMAND
FORMER UK SECURITY AND INTELLIGENCE
COORDINATOR AND MEMBER OF PALADIN
STRATEGIC ADVISORY GROUP



INTERNATIONAL INVESTMENT CASE STUDIES

The firms outlined below are reflective of current innovation within the SafetyTech sector. They showcase the growing need within the market for services that support safety by design development, and the need to monitor and guard against both physical threats and those which exist online. Please note the examples are sourced via company web data and are considered accurate at the time of review.

FIRMS WHICH SUPPORT INTELLIGENCE BUILDING AND GUARD AGAINST DISINFORMATION	
 <p>Company Name: Factmata</p> <p>Key Services: Intelligence Reporting, Brand Management</p> <p>Founded: 2017</p> <p>Investment: \$4m</p> <p>Based: UK / US</p>	<p>Established in 2017, Factmata aims to promote a better understanding of content through its online scoring system. It takes online content into consideration, and once it reads it intelligently, it assigns it a "Trust Score." This score is used to support critical thinking by users. Work is ongoing at Factmata to support integration with browser extensions, community platforms, apps and ad-blocker services.</p> <p>Factmata uses advanced natural language processing that reviews content and learns what is typically deceptive content. The AI is supported by a team of journalists, researchers and advocacy groups who also provide feedback to improve services.</p> <p>Factmata currently offers intelligence report, moderation and API services, with application across PR agencies, government, brands and digital platforms. The project has won multiple awards, including the UNESCO Netexplo Award (2019) and the CogX London 2018 award for Best API product in politics and has received funding from both Horizon 2020 European Framework for Research and Development and the UK Government.</p> <p>Factmata has also received funding from multiple private funders, raising over \$4m from investors that include search engine Google, European seed investor Seedcamp, Adblock Plus parent company Eyeo, pre-seed and seed-stage investor Dig Ventures, early-stage venture firm Fine Day Ventures, and a range of private investors, including Twitter's Biz Stone.</p>
 <p>Company Name: Nisos Inc.</p> <p>Key Services: Cyber Security and Managed Intelligence</p> <p>Founded: 2015</p> <p>Investment: \$19.2m</p> <p>Based: US</p>	<p>Nisos is the Managed Intelligence company. Their services enable security, intelligence, and trust & safety teams to leverage a world-class intelligence capability tailored to client needs. They fuse robust data collection with a deep understanding of the adversarial mindset delivering smarter defense and more effective response against advanced cyber-attacks, disinformation and abuse of digital platforms.</p> <p>With the increasingly sophisticated threat actors in today's society, Nisos co-founders took their collective experience working with the US Government and developed an array of solutions to support firms as they address cybersecurity risks faced globally.</p> <p>Formed in 2015 and operating as a trusted partner to the Fortune 500, Nisos aims to address core issues faced by firms such as disinformation, platform abuse, and nation-state level hacking. It does this through its offering of context-based research, and action-orientated monitoring of client platforms, which can determine the existence of threats and plan how best to address. Nisos supports firms in combating disinformation, and in monitoring and promoting platform integrity. Nisos also helps companies build their threat programmes, and support mergers, acquisitions and third-party due diligence checks.</p> <p>Since 2015 Nisos has received almost \$20m from investors. Nisos recently announced a \$6m funding round led by global cyber investor Paladin Capital Group. Existing investors Columbia Capital and Skylab Capital also participated in the round. The investment enables Nisos to expand its marketing and operations, while also extending its international footprint.</p>



Company Name:
Graphika

Key Services:
**Network and
Cluster Analysis**

Founded:
2013

Investment:
\$6.5m

Based:
US

Graphika is a network analysis firm, which maps and clusters rapidly growing weblog communities using graph analysis and machine learning. Graphika has a strong affiliation with Harvard's Berkman Klein Center, its services have been used to map online networks for an assortment of academic, not-for-profit, and commercial clients – discovering the contours of “cyber-social terrain.” Graphika was initially used in publishing and is rapidly expanding its commercial base to include automotive, health and medicine, sports, and global consumer brands.

Key services offered by Graphika include disinformation detection, analysis and monitoring, target audience discovery and trend monitoring, campaign optimization and impact measurement, influencer discovery and audit, CEO and executive leadership, influence analysis, strategic communication and crisis response.

Graphika has raised over \$6m in funding since 2014. This funding was provided in part by Lavrock Ventures, a venture capital firm specializing in early-stage (Series A) investments, seeking to invest in cybersecurity and enterprise software companies. Their investments are often in commercial-focused technologies with potential relevance to National Security.

In more recent years, Graphika has started to work with social media firms to identify patterns in harmful behavior. In 2018 the firm helped identify an overlap in social media users who supported President Trump and engagement with QAnon material. More recently, work with Facebook has identified and supported the takedown of reported China-based campaigns put in place to further polarize American politics during the recent election campaign.

FIRMS WHICH OFFER SAFETY BY DESIGN AND MONITORING SERVICES



Company Name:
SuperAwesome

Key Services:
**Safety by Design and
Content Moderation**

Founded:
2013

Investment:
\$73m

Based:
UK / US

SuperAwesome's award-winning technology provides the tools for safe digital engagement with almost half a billion kids every month. Coming from a background in digital media, technology, and kids' media, SuperAwesome was founded in 2013 as a response to the increased number of devices, screens, channels, sites and apps available to young people.

Since 2013 SuperAwesome has developed kidtech technology, which offers strict compliance and functionality, and has grown to become a top supplier, supporting over 300 of the major global brands and content creators.

Services offered by SuperAwesome include: AwesomeAds, which is the only ad platform built for the global kids industry; SuperAwesome Creators, which supports ethical influencing and engagement with under 16s and families; Kidsafe Social Video, which offers Kid-safe, brandsafe, contextual advertising on YouTube and other streaming platforms; Kids Web Services, which helps web developers build personalized digital experience for kids; and PopJam, which is an under-13s social engagement platform.

Since 2013 SuperAwesome has received \$76m in investment, with the most recent being a Series C investment. In September 2020 SuperAwesome was acquired by Epic Games, who are known for the development of Unreal Engine and the successful Fortnite game, which has 350m users. The acquisition by Epic Games³⁸ ensures that the firm can deliver its promise of a safe, privacy-driven digital experience at scale.

³⁸ SuperAwesome (2020) 'SuperAwesome joins the Epic Games family' Available at: <https://www.superawesome.com/superawesome-joins-the-epic-games-family/>



Company Name:
OpenSlate

Key Services:
**Advertising and
Content Moderation**

Founded:
2012

Investment:
\$18m

Based:
US

Founded in 2012 and headquartered in New York, OpenSlate is one of the only companies that can comprehensively measure brand safety, suitability and context for advertisers. Ratings generated by OpenSlate power billions in video spend from more than 800 clients. The service is in use across 37 international markets, and OpenSlate's innovative data science and independent content ratings are used to better understand and assess content quality, safety, suitability, and subject matter – which ultimately supports brands in making better decisions about where their advertising runs.

OpenSlate technology is in use by some of the world's largest digital video platforms, including YouTube, Facebook and from October 2020, TikTok. With over 100m monthly US users, OpenSlate's new deal with TikTok reflects the growth in conscientious advertising and increased consideration of online branding.

The new TikTok Brand Safety Solution works to filter out inappropriate content and provide brands with confidence that their products are being advertised in a suitable context. TikTok's partnership with OpenSlate came as a reaction to growing concerns among brands, who saw the clear risks of advertising on such a fast growing site.

FIRMS OFFERING INTELLIGENT SURVEILLANCE SERVICES



Company Name:
X.Labs

Key Services:
**Threat detection,
protection and
digital healthcare**

Founded:
2018

Based:
US

X.labs is a leader in next generation safety threat detection and digital healthcare technologies with the overarching aim of preserving life. The company's products are varied and include:

- X1 and XLe: A connected IoT device that uses X.Labs proprietary software algorithms, hardware interface, artificial intelligence and best-in-class sensor technology to detect and mitigate threats in public spaces.
- PreCheck: A digital healthcare product that includes a Bluetooth thermometer and healthcare app allowing for at-home COVID-19 testing.
- Feevr: Feevr is an AI based system for screening and detecting people in crowds with an elevated temperature.
- ShoX: Anchored on the Blockchain, ShoX is a breakthrough in less-than-lethal ballistic products for law enforcement, military and anti-terrorism operations.

The innovation undertaken at X.labs reflects the potential of safety technology and its potential for addressing issues such as COVID-19 and gun violence. Products such as the X1 and XLe could potentially play a role in reducing gun violence and subsequent injury related to this, and products such as Feevr illustrate the firm's capability to innovate and meet markets needs.



Company Name:
Viisights

Key Services:
Surveillance
and detection

Founded:
2015

Investment:
\$11.2m

Based:
Israel / US / Singapore

Viisights is a behavioral recognition system for real-time video intelligence founded in 2015 with offices in the US, Israel and Singapore. Viisights' video understanding technology is based on a unique implementation of deep neural networks. These networks are capable of analyzing and deducting high level concepts derived from video content. Viisights technology recognizes the behavior of diverse objects, as well as its relevant contexts and can therefore differentiate what is considered deviant behavior based on location. The technology has been integrated at scale in Leon, Mexico and used by police to identify deviant behavior, support perimeter protection, and monitor environmental changes such as smoke or fire.

Viisights has three core products: viisights wise which supports widespread surveillance cameras located throughout traffic control centers, roads, public areas, buildings, shopping centers, commercial and industrial zones; viisight true which enhances the control of security guards and reduced false alarms; and in-cabin monitoring which supports the monitoring of activity in rental cars.

Since 2016 Viisights has received \$11.2m in investment. This investment was provided by AltaIR Capital, ASF Ventures, Leta Capital, Maxfield Capital, Firsttime and TheTime.



Market Trends & Opportunities

There are a wide range of driving factors behind the growth of investment and activity in Safety Tech provision internationally. This includes:

- **Increased consumer awareness regarding online harms, alongside an expressed loss in trust of many consumers within tech platforms:**

In recent years, citizens have become more knowledgeable regarding their right to privacy, whilst ensuring safety on platforms. This has meant that online safety is much higher on consumer agendas.

- **Market preference to minimize toxicity and abuse from platforms, and to disassociate platforms from harmful or damaging material:**

As reflected by Epic Games acquisition of SuperAwesome, we perceive that gaming and online platforms are becoming more cognizant of online harms, and the need to use enhanced technology to address the problem space. For example, if an online game is poorly moderated, this can result in loss of community and therefore, a loss of associated revenue.

- **The increasing volume of content online requires a move towards using AI in the content moderation process:**

There is a consensus that online content is now created faster and in greater volumes than can currently be solely screened or moderated by humans. Using new solutions such as AI and natural language processing (NLP) can increasingly help human moderators better detect harmful content, material, or sentiment in real-time, and reduce moderators' own exposure to harmful content.

- **Heightened concerns regarding the spread of disinformation in the public domain, and ensuring election integrity:**

Disinformation is a commonly used tool online, often deployed by sophisticated threat actors to generate false narratives. Countering misinformation and disinformation and providing transparency in information reporting is critical to a democratic society, e.g., through providing live fact-checks, and blocking or removing deliberately misleading content from online platforms.

- **Increased usage of technology to identify risk in the public domain**

emerging online technologies reflect the symbiotic relationship between the real-world and cyberspace.

For example, the use of computer vision technologies to identify targeted violence and terrorism (TVT) physical threats such as weaponry in public spaces and notify law enforcement.

- **Safety Tech provides new opportunities for investment in technologies that deliver on Environmental Social and Governance (ESG) criteria,**

particularly when cyberspace is conceptualized as an environment. Online Safety Technologies seek to generate financial returns while also creating a positive social and online environmental impact; therefore, Safety Tech meets the criteria for 'Impact Investing' and what can be now conceptualized as 'Cyber CSR.'



International Comparison

The map sets out an overview of some key regions identified in Safety Tech research and market analysis to date.

CANADA

Two Hat: Two Hat Security is dedicated to keeping online communities safe from harassment and abuse. Two Hat provides an AI-powered content moderation platform that detects high-risk content for online games, social networks, messaging apps, and more.

DENMARK

SafeonNet provides online reputation insurance and protection. This includes support with removing unwanted content online.

NETHERLANDS

Deep Trace Labs: provides a 'deepfakes' detection solution designed to ensure the integrity of visual media.

GERMANY

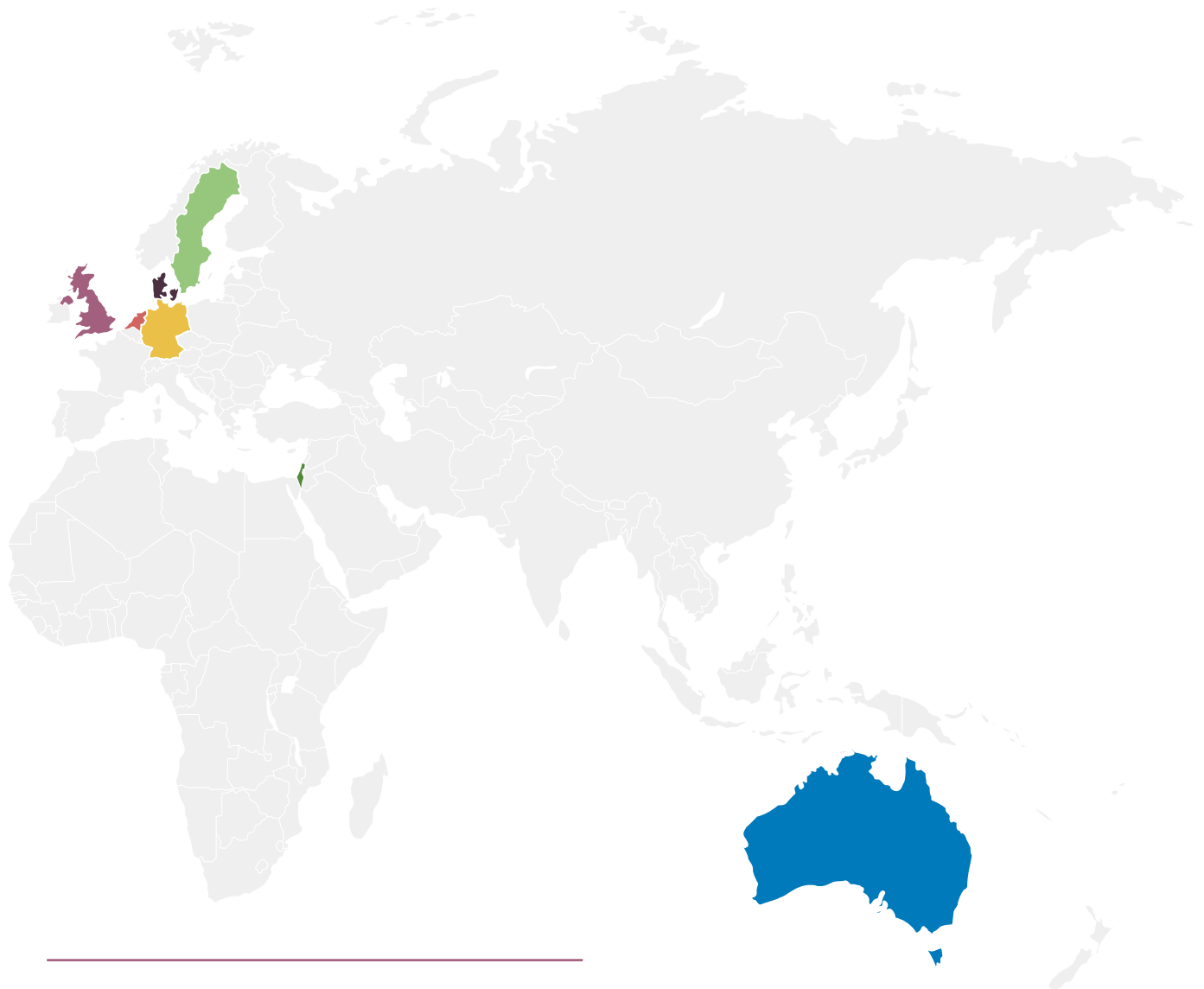
Germany adopted its Network Enforcement Act ('NetzDG') in 2017. This law requires online platforms with more than two million registered users in Germany to remove 'manifestly unlawful' content – non-compliance risks a fine of up to €50 million. They are now recognised as an emerging growth market for Safety Tech.³⁹

SWEDEN

Besedo provides content moderation services, including profanity filters. AI-powered nudity detection, and customized AI moderation models.

NetClean are experts in detecting child sexual abuse material in business IT environments through using hashing technology.

³⁹ Safety Tech Innovation Network (2021) 'German Safety Tech industry gains momentum' Available at: <https://www.safetytechnetwork.org.uk/articles/german-safety-tech-industry-gains-momentum>



UNITED KINGDOM

In May 2021, the government published the Online Safety Bill. This Bill sets out a new regulatory framework for online safety and companies' responsibilities to keep UK users, particularly children, safer online - including robust action to counter illegal content and activity.

The UK is also recognised as a world-leader in Safety Tech and has committed funding to support the sector develop innovative new solutions and protect as many users as possible globally.

ISRAEL

Jiminy provides parents and guardians with the information they need to understand what their child is doing online and will provide an alert if children are subject to social issues, concerning content, or toxic device usage.

AUSTRALIA

Australia established an eSafety Commissioner through its Enhancing Online Safety for Children Act in 2015. The eSafety Commissioner is responsible for promoting online safety for all Australians. It developed a Safety by Design initiative in 2018, which places user safety at the forefront of online service development — creating safer environments online. In June 2021 the Australian eSafety Commissioner launched a range of interactive assessment tools to be utilized for free by the region's tech companies. "Established in line with its Safety by Design initiative, use of the tools ensures that user safety remains front and center of tech product design and development. These free tools are a critical step towards bolstering the global effort to make people safer online."⁴⁰

⁴⁰ Fintech Times (June 2021) 'Australian Internet Safety Regulator Launches Free Assessment Tools to Improve Online Safety' Available at: <https://thefintechtimes.com/australian-internet-safety-regulator-launches-free-assessment-tools-to-improve-online-safety/>

Recommendations



There is a need to create a road map for how the US Safety Tech sector may be supported to grow sustainably, and benefit as many users as possible. We would like to encourage stakeholders to connect with our team at Paladin to build the partnerships critical for success in the new Safety Tech industry.

Our recommendations for growth of the emerging Safety Tech Sector are as follows:

- 1 Promote and raise awareness of the US Safety Tech sector, the innovative providers of tech solutions to tech facilitated criminal and harmful behaviors.
- 2 Showcase US Safety Tech providers, create stronger networks to connect buyers and suppliers, to drive adoption and grow exports.
- 3 Support Safety Tech firms to access the right forms of capital for growth.
- 4 Conceptualize how cybersecurity and Safety Tech can work in conjunction to optimize security and safety in cyber contexts.
- 5 Invest in the usage of online safety technologies to identify risk in the public domain specifically in the area of targeted violence and terrorism.
- 6 Encourage public policy action and private investment in Safety Tech, specifically regarding impact investing, ESG criteria and Cyber CSR.
- 7 Consider the policy and regulatory environment – with government providing leadership, guidance and appropriate legislation to address online harms.
- 8 Support the development of collaborative partnerships to promote sharing of data, information and best practices in tackling online harms.
- 9 Design and implement a sector-wide Safety Tech strategy to build an ecosystem; including initiatives to help cross-sector collaboration, such as, government and industry roundtables; innovation networks and development of a US Safety Tech Industry Association.
- 10 Establish public and private partnerships to drive innovative solutions, cross-disciplinary research and development in Safety Tech.

Appendix



APPENDIX A

Research Team and Acknowledgements

Principal Investigator: Mary Aiken PhD is a Professor of Cyberpsychology and Chair of the Department of Cyberpsychology at Capitol Technology University. She is a Professor of Forensic Cyberpsychology in the Department of Business and Law at the University of East London (UEL), Adjunct Professor at the Geary Institute for Public Policy University College Dublin (UCD). Member of the INTERPOL Global Cybercrime Expert Group, Academic Advisor to Europol's European Cyber Crime Centre (EC3), Fellow of the Royal Society of Medicine, member of the Medico-Legal Society, International Affiliate Member of the American Psychological Association (APA) and Fellow of the Society for Chartered IT Professionals. Professor Aiken was study Co-Lead for this research she is a member of the Paladin Capital Strategic Advisory Group.

Perspective Economics

(www.perspectiveeconomics.com) provides independent economic analysis and advisory services. They use modern methodologies and analytics to gather and interpret evidence, providing insight and clarity on important economic, business and social issues. They have significant experience in conducting sector studies across the UK, exploring the economic contribution and labor markets of emerging industries and sectors such as cybersecurity, HealthTech, AI, and Safety Tech. The study Co-Lead for this research was Sam Donaldson (Director).

Paladin Capital Group is a leading global investor that supports and grows the world's most innovative companies through venture investment, expansion, and growth capital. We are headquartered in Washington, DC, with offices in New York City, Silicon Valley, and London. We seek out diverse investment opportunities across the US and worldwide. Paladin is a leader in investing in technologies, products, and services focused on dual use in both commercial and governmental markets with a strong value-add culture.

We would like to thank the UK's Department for Digital, Culture, Media & Sport (DCMS) for their work to date in helping to establish the global Online Safety Technology ecosystem.

Funding: this research report was funded by the Paladin Capital Group.

APPENDIX B

Methodology

This methodology is consistent with the UK's 'Safer Technology, Safer Users' research to enable comparisons to be drawn between the relative strengths and opportunities for the sector.

METHODOLOGY	DESCRIPTION
Desk Research	<p>Using a Grounded Theory⁴¹ approach, the team reviewed over eighty pieces of academic literature, sector overviews, and grey literature relating to online harms, and online safety technology globally.</p> <p>We longlisted more than 500 potential Safety Tech businesses within the United States and internationally, to identify the characteristics and offer of Safety Tech organizations, prior to shortlisting to identify a final list of Safety Tech firms operating in the United States.</p> <p>This was undertaken through identifying a series of keywords and terms extracted from known Safety Tech providers, products and solutions (e.g., moderation, filtering, abuse, misinformation etc.) and synonymous terms.</p> <p>These key terms were matched against investment platforms such as Crunchbase and Pitchbook, as well as web data and LinkedIn company descriptions. This enabled the team to build a long-list of potential Safety Tech businesses within the United States. Each business was subject to automated and manual review and assigned against each area of the Safety Tech taxonomy (or removed from the analysis if considered less applicable to the research).</p> <p>This yielded the shortlist of firms identified for this study.</p>
Definition and Market Scoping	<p>The research team developed a working definition for what constitutes 'Online Safety Tech'. This also informed building on the 'Safety Tech' taxonomy within this research, and the short-listing of Safety Tech firms, i.e., those that provide a product or service aligned to the categories.</p>
Market Analysis	<p>We identified relevant companies and trading information for the dedicated Safety Tech providers using Crunchbase, Pitchbook, and company websites. Investment data has been sourced from Pitchbook.</p>

Given the emergence of the Safety Tech sector, data is limited in a number of areas, and will develop as the sector grows. Therefore, all references to market size, employment, and investment within this report are estimates only, using bottom-up modelling and direct stakeholder research.

⁴¹ Glaser, B., Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.



Perspective
Economics



PALADIN
CAPITAL GROUP

Paladin Capital Group

With offices in locations around the world, Paladin is a leading global investor.

WASHINGTON DC OFFICE

2020 K Street NW, Suite 620
Washington, DC 20006
202-293-5590

LUXEMBOURG OFFICE

43 avenue John F. Kennedy
L-1855 Luxembourg
+352 26 00 5758

NEW YORK CITY OFFICE

295 Madison Avenue
12th Floor
New York, New York 10017
202-293-5590

MENLO PARK OFFICE

3000 Sand Hill Road
Suite 2-145
West Menlo Park, CA 94025

LONDON OFFICE

20 North Audley Street
London, UK W1K 6WE
+44 (0)203 931-9704

#SAFETYTECH