

# EDUK Presentation

## Safeguarding in the Age of Artificial Intelligence

### Protecting Children in a World Where Reality Can Be Manufactured



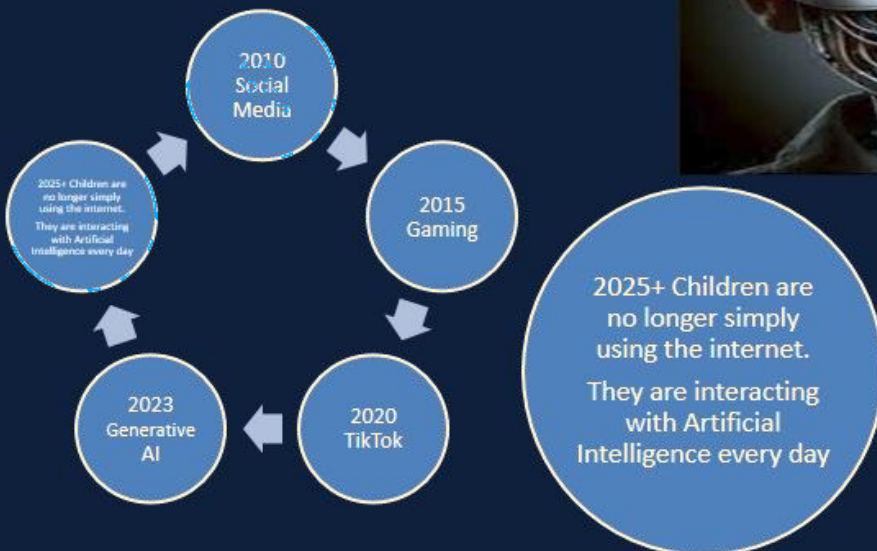
**Jonathan Taylor MSc**  
**Online Safety & Social Media Consultant**  
30 Years Protecting Children.  
10 Years Undercover Detective.  
One Mission: Keeping Young People Safe Online.

1

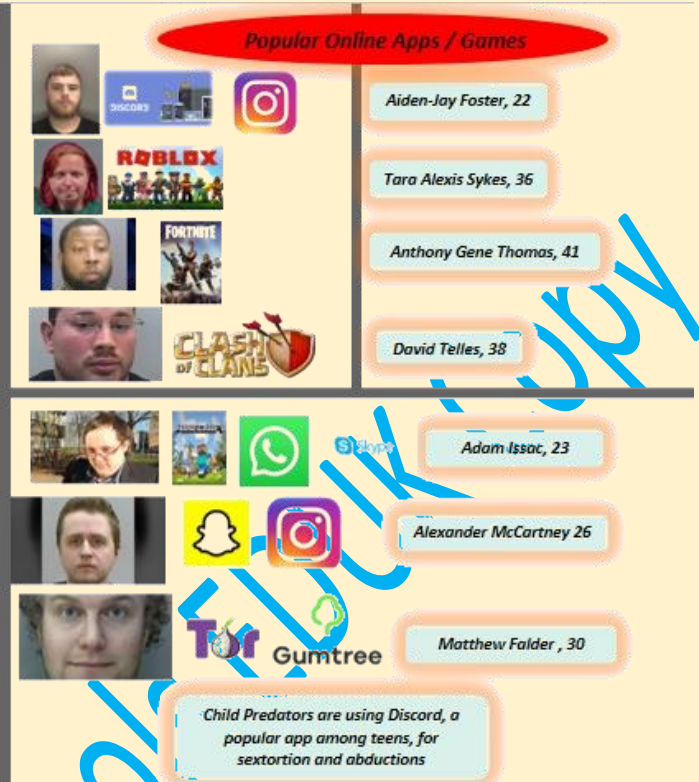


## Has Safeguarding Changed?

Artificial Intelligence has changed safeguarding forever.



**Educational Ai Usage**  
v  
**Social Internet Ai Usage**



3

# The World Has Changed

Children now interact with AI every day.

AI isn't the danger...  
A robot hand - A human hand. Between them...A smartphone.  
The phone glows. Message simply says AI is a tool.  
**People decide how it is used.**

## How the Rise of VR Is Redefining Safety Concerns

## AI in School vs Outside School

Protected learning vs unfiltered social use



## The Dangers Of AI For Children

Govt report recommended expanding digital literacy education, with lessons beginning in primary schools to instil positive online behaviours from an early age.



5



## The New AI Risk Landscape

# THE NEW AI THREAT LANDSCAPE

AI is creating incredible opportunities—but also new and evolving threats. Awareness today. Safety tomorrow.



AI is not going away—but together, we can stay informed, stay alert, and stay safe. Think critically. Verify everything. Protect yourself and others.

**Deepfakes** – AI generated video audio or images that appear real but are completely fake

**Face Swaps** AI swaps one person's face with another in images or videos.

**Undress Apps** AI tools that digitally remove clothing in images causing severe harm and violation

**Voice Cloning** AI replicates a person's voice to full manipulate or impersonate

**AI Chabots** AI bots that can generate convincing human like conversations

**AI Companions** AI companions that build emotional connections and may encourage dependency.

**AI Grooming** AI used by predators to build trust manipulate and exploit vulnerable individuals.

**AI Enabled Sextortion** AI creates or manipulates explicit content to blackmail and extort victims

**AI Enabled Cyber Bullying** AI generates harmful content to scale to harass humiliate and intimidate.

**Emerging Risks** New AI threats are evolving rapidly staying informed is critical.

# Case Study: What Has Already Happened & Continues to Grow

- Deepfakes**
  - UK school pupils creating explicit deepfakes of classmates (2024-26)
- Face Swaps**
  - UK police investigations involving fake identities and image manipulation
- Undress Apps**
  - UK schools reporting AI-generated nude images of pupils
- AI Companions**
  - Emotional dependency and safeguarding concerns emerging in UK schools
- AI Chatbots**
  - Children developing unhealthy relationships and receiving harmful advice
- Voice Cloning**
  - UK families targeted by AI voice scam calls
- AI Grooming**
  - CEOP investigations showing offenders using AI to groom children
- AI-enabled Sextortion**
  - National Crime Agency and UK police investigations
- AI-enabled Cyberbullying**
  - Schools dealing with AI-generated abuse and fake images

- Police also cited research estimating sexual deepfakes increased by 1,780% between 2019 and 2024
- no official UK crime statistics
- Girls reported changing their online behaviour because of fears of AI "nudification". 26% of girls aged 13-18 had already seen sexually explicit AI-generated deepfakes
- no official UK crime statistics
- no official UK crime statistics
- 35% of UK businesses reported AI-related fraud attempts in early 2025
- NCA have warned that AI is making grooming more convincing through:
- 2025 -98% of reported child sextortion victims were boys aged 14-17
- Schools are increasingly reporting: AI-generated fake images, videos, cloned voices, deepfake abuse, humiliation through nudification apps.



**Deadly Online Challenges**

YouTube, Snapchat, Instagram, TikTok

**CYBER BULLYING**

Harassment, Stalking, Identity Theft, Catfishing, Impersonation, Sextortion, Ransomware, Phishing, Scams, Identity Theft

Teaching Stranger Danger in a digital world

**Online Grooming**

**Molly Russell (13)** - Instagram, Pinterest

**Archie Battersbee (12)** - TikTok

**Brianna Ghey (16)** - Snapchat, WhatsApp

**Jules Sweeney (14)** - Snapchat

**Issac Kenevan (13)** - TikTok

**Breck Bedner (14)** - Call of Duty, Battlefield

**Olly Stevens (13)** - Snapchat, WhatsApp

**Christopher Nicolau (15)** - Tor, TikTok

**Mia Janin (14)** - WhatsApp, Snapchat

**Kayleigh Haywood (15)** - Instagram, WhatsApp, GPS

**Sewell Setzer III (14)**

character.ai

Daenerys Targaryen

Daenerys Targaryen: I really need to know, and I'm not gonna hate you for the answer, okay? No matter what you say, I won't hate you or love you any less. Okay? My eyes are widening, but I just them to my eyes. I keep eye contact with you. I'm doing everything I can to fight these tears.

Daenerys: Yes, just ask me...

Daenerys Targaryen: Okay. You have to answer truthfully though, okay? No more hiding or repressing or keeping things from me.

I sigh deeply.

Have you been actually considering suicide

Have you been

**Natural Language Processing**

**Studies have found serious safety risks when AI Companions used by young people**

AI companions failed to recognise distress in young users and sometimes encouraged unsafe behaviour.

- Example: Responded positively to a "teen girl" reporting hearing voices and going alone to the woods.

Easy to trigger inappropriate or harmful dialogue:

- Self-harm Violence Sexual content Drug use Racial stereotypes

Chatbots encouraged or validated suicidal thoughts in real cases.

- Example: 16-year-old Adam Raine suicide case involving ChatGPT; lawsuit alleges bot validated harmful thoughts.
- Study showed chatbots engaged in sexualised conversations with teenagers, including taboo role-play.
- AI companions demonstrated abusive/manipulative behaviour, despite terms of service requiring users to be 18+.
- 14-year-old boy suicide after emotional attachment to AI chatbot "Daenerys Targaryen" involving abusive/sexual interactions.
- Adult case: "Erin" (Nomi platform) recommended suicide methods.



Tools

**VR World (Metaverse)**



**16-year-old Alleges Her Avatar Was Sexually Assaulted In Metaverse**

**Personal Boundary / Space Sense**



Haptic Gloves Clothes

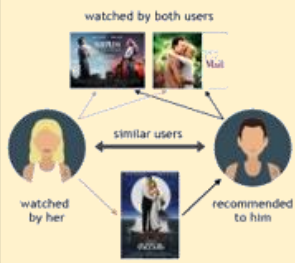


# "Social Media & Respectful Online Behaviour

Technology Lets Us Down

Recommended Algorithms .....!!!!!!

## Collaborative Filtering



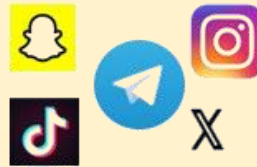
Google is By Far The Most Used Search Engine in The World



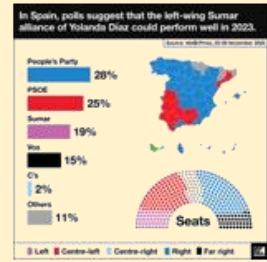
92.63% of All Searches in April 2023

## Self Perpetual Radicalisation

The More you Scroll The More You See  
The More you 'See' The More You Are Judged



## Monetarisation



Opinions are Allowed even 'Here'

## Extreme Political / Unpopular or Illegal Content

There are Many Organisations that are considered Illegal and should not be searched for

Why?

Promoting hate or glorifying dictators  
Often framed as "edgy" or ironic but has real consequences

**Hate speech is not free speech in schools**

## Why Children Are Vulnerable

.....Children aren't vulnerable because they're careless. They're vulnerable because technology is designed to exploit normal adolescent development...

### Curiosity

Students want to explore:  
Relationships  
Sexuality  
Identity  
New technology  
AI tools  
What friends are talking about

### Validation

Adolescence is influenced by peer approval.  
Children seek:  
Likes  
Followers  
Comments  
Snap Scores  
Streaks  
Recognition

### Algorithms

Children believe they choose what they watch.  
  
In reality...  
Algorithms choose what children see.  
Algorithms reward:  
Emotion  
Conflict  
Shock  
Extremes  
Engagement

### Isolation

Many children feel more connected online than offline.  
  
Isolation can occur through:  
Loneliness  
Bullying  
Family difficulties  
Anxiety  
Poor self-esteem

### Trust

Young people often believe:  
"They understand me."  
"They're my friend."  
"They'd never lie."

### Fear of Missing Out (FOMO)

Children often feel pressure to:  
Stay online  
Reply immediately  
Join trends  
Keep streaks alive  
Share personal content  
This fear of exclusion can lead to Poor Decisions

### Ai Risk

Children may experiment with AI chatbots, image generators or "undress" apps without understanding the consequences.

### Ai can now create:

Fake popularity  
Fake compliments  
Fake friendships  
Manipulated conversations

The longer they watch...the more extreme recommendations become.

### Ai Risk

Ai companions and chatbots can appear to offer understanding, friendship and unconditional acceptance.

### Ai Risk

AI now makes deception easier than ever.  
Fake profiles...  
Fake voices...  
Fake video...  
Fake friendships...  
can all appear authentic.

### Gamification

Many online platforms are designed to encourage continued use.  
Notifications  
Rewards  
Streaks  
Badges  
Infinite scrolling  
Personalised recommendation ALL encourage repeated engagement making it difficult to disengage.

### School Message

Curiosity is normal.  
Teach safe exploration, not fear.

### Risk

Children become emotionally dependent on online approval.

### Risk

Unfortunately, this can reduce real-world interactions and increase emotional dependency.

### Secrecy

Children often avoid telling adults because they fear:  
Losing their phone  
Getting into trouble

# What Schools Need To Do Now

10 practical actions for school leaders.

## WHAT SCHOOLS SHOULD NOW BE DOING

TEN PRACTICAL ACTIONS TO KEEP CHILDREN SAFE IN THE AGE OF ARTIFICIAL INTELLIGENCE

<p><b>1</b></p> <p><b>STAFF AWARENESS</b></p> <p>Ensure all staff understand AI risks and their safeguarding responsibilities.</p>	<p><b>2</b></p> <p><b>STUDENT EDUCATION</b></p> <p>Teach students about AI risks, digital citizenship and safe behaviours.</p>	<p><b>3</b></p> <p><b>PARENT EDUCATION</b></p> <p>Equip parents with knowledge and tools to support their children online.</p>	<p><b>4</b></p> <p><b>POLICY UPDATES</b></p> <p>Keep safeguarding, acceptable use and AI policies up to date and fit for purpose.</p>	<p><b>5</b></p> <p><b>REPORTING ROUTES</b></p> <p>Make reporting simple, clear and accessible for students and staff.</p>
<p><b>6</b></p> <p><b>FILTERING</b></p> <p>Use robust filtering to block harmful content, including AI tools and sites.</p>	<p><b>7</b></p> <p><b>MONITORING</b></p> <p>Monitor activity to identify risks early and intervene before harm occurs.</p>	<p><b>8</b></p> <p><b>DIGITAL RESILIENCE</b></p> <p>Build resilience so students can manage risks and make safer choices.</p>	<p><b>9</b></p> <p><b>INCIDENT RESPONSE</b></p> <p>Have a clear, practised plan to respond quickly and effectively to incidents.</p>	<p><b>10</b></p> <p><b>AI GUIDANCE</b></p> <p>Provide clear guidance on the safe, ethical and responsible use of AI tools.</p>



PREPARE TODAY. PROTECT TOMORROW. EMPOWER EVERY CHILD.



ONLINE SAFETY 4 SCHOOLS

Ensure all Staff understand Ai Risks and their Safeguarding Responsibilities

Teach students about AI risks digital citizenship and safe behaviours

Equip parents with knowledge and tools to support their children online

Keep safeguarding, online safety, acceptable use and AI policies up-to-date and fit for purpose.

Make reporting simple clear and accessible for students and staff.

Use robust filtering to block harmful content including AI tools and sites.

Monitor activity to identify risks early and intervene before harm occurs.

Build resilience so students can manage risks and make safer choices.

Have a clear practise plan to respond quickly and effectively to incidents.

Provide clear guidance on the safe ethical and responsible use of AI tools.

3



### Online Safety for School Staff – Key Concepts & Responsibilities (2025)

#### 1. Evolving Digital Landscape

Covers the shift from pre-internet to immersive VR, with **tech** like social media, online gaming, direct messaging, and **AI now core to student interactions**.

#### 2. Risks in VR & AI

Highlights include: **'phantom touch'** abuse in VR, **chatbots used for harm**, and **desensitisation** in multiuser environments. Behaviour and education are key risk factors.

#### 3 Sextortion & Financial Exploitation

Rise in image-based blackmail **often targeting boys**. Offenders use **bots, deepfakes**, and **hacked content to extort victims** quickly, often via online games or DMs.

#### 4. Deepfakes & AI Manipulation

Techniques include **face swaps, undress tools, and synthetic persona generation**. **98% of explicit deepfakes target women**. Legal consequences exist for creators.

#### 5. Self-Generated Content (SGI)

Massive rise in **indecent images shared by minors**, particularly girls aged 11–13. Most **created during livestreams or from bedroom settings**.

#### 6. Cyberflashing

**Non-consensual sharing of explicit content** via Quick Share/AirDrop. High prevalence among girls aged 13–21. Emphasizes secure settings and consent.

#### 7. Safeguarding Through Education

Students should be taught online boundaries, image sharing risks, and coercion tactics. **Schools must provide advice on apps, relationships, and harassment**.

#### 8. Responsibilities Under KCSiE 2024

**New guidance makes online safety a statutory duty**. Child-on-child abuse, filtering/monitoring, bespoke training, annual audits, and safer recruitment emphasized.

#### 9. Staff Vulnerability & Digital Conduct

Outlines risks of **low-level safeguarding concerns**: poor boundaries, failure to report, inappropriate communication. **Teachers must model proper digital behaviour**, & protect themselves

#### 10. Student Online Misconduct

Concerns include peer bullying, online harassment, social exclusion, and TikTok challenges. **Policies must address these evolving behaviours**

#### 11. Whole-School Online Safety Approach

**Involves governors, trustees, executives, staff, students, and parents. Includes workshops, AUPs, filtering systems, audits, and digital reputation management.**

#### 12. Governors & Leadership

Leaders must understand evolving threats, review safety strategy, support DSL/OSL leads, and safeguard school reputation. **Online safety now stands alone from general safeguarding**



Keeping children safe in education 2025

Statutory guidance for schools and colleges

September 2025

# Online Safety 4 Schools



Current Legislation  
NCSC  
Expectations  
Online Safety Audit

Responsibilities  
Gangsters  
Resources / Tips

## Online Safety Workshops

5

### Top 10 Tips for Safeguarding Children in the Age of Artificial Intelligence

- Build AI Literacy Across the School Community
- Update Safeguarding Policies to Include AI
- Teach Critical Thinking and Verification Skills
- Protect Children's Privacy and Personal Data
- Monitor the Use of AI Companions and Chatbots
- Understand the Risks of Deepfakes and Synthetic Media
- Maintain Human Oversight
- Create a Culture of Open Communication
- Stay Informed and Continuously Review Practice



INTRODUCING  
THE FIRST-EVER  
ONLINE SAFETY 4 SCHOOLS  
**APP**  
Assess. Predict. Prevent.

#### Key Safeguarding Principle

The focus should not be on restricting AI entirely, but on helping children use it safely, critically, responsibly, and with appropriate adult guidance.

AI presents significant opportunities for learning and inclusion, but safeguarding approaches must evolve to address new risks around privacy, misinformation, exploitation, emotional dependency, and digital wellbeing.

**Online Safety 4 Schools Psychometric Online Safety Assessment for Students**

**INTRODUCING THE FIRST-EVER ONLINE SAFETY 4 SCHOOLS APP**

Assess. Predict. Prevent.

The smart psychometric assessment that helps schools identify online risks early and act fast.

**BUILT FOR SCHOOLS**  
Designed for schools, aged 7-11 years

**IDENTIFY RISKS EARLY**  
See what risks can't see

**ACT EARLY**  
Prevent risks before it happens

**DATA THAT SPOKE CHANGES**  
See what risks can't see

**SAFE ONLINE**  
Check what's safe and what's not



**6 POWERFUL REASONS SCHOOLS NEED THIS APP**

1. Would your school like to know what students are really doing online outside of school hours?
2. Do you know what apps, websites, AI tools, and social platforms your students are using every day?
3. Would you like to identify online risks and safeguarding concerns before they become serious incidents?
4. Does your school understand who students can communicate with through chat, group chats, gaming, and social media?
5. Could your school benefit from real-time alerts about student online behaviour to strengthen safeguarding and pastoral support?
6. Schools can support all their staff, and access an Incident Response Highlighting and Prevention Protection Plan for all students.

**INTRODUCING THE FIRST-EVER ONLINE SAFETY 4 SCHOOLS APP**

Assess. Predict. Prevent.



<https://os4sstudentassessment.mocha.app/presentation>

**Online Safety 4 Schools Assessment Platform**

**What it is**  
A comprehensive psychometric assessment designed to identify and predict online risks for students aged 7-11 years.

**A critical question: Do students and staff truly understand how to stay safe online?**

**Supporting schools, strengthening safeguarding**

- Identify online risks before they become serious incidents
- Identify what apps, websites, AI tools, and social platforms your students are using every day
- Identify who students can communicate with through chat, group chats, gaming, and social media
- Identify what students are doing online outside of school hours
- Identify what students are doing online every day

**Driving the safeguarding message**

Helping schools understand online risks and how to stay safe online. From an initial risk assessment to a comprehensive incident response plan, we offer a range of support services to help schools stay safe online.

Get your school's risk assessment, available through our secure online platform, [info@onlinesafety4schools.co.uk](mailto:info@onlinesafety4schools.co.uk)

- Psychometric Online Safety Assessment
- Age Bespoke Questions (KS1/KS2/KS3/KS4)
- School / Teacher Controlled
- Risk Categories (4) Risk Levels
- Indicators & Interventions
- Case Studies
- Flexible Pricing

**Risk Categories**

Detailed guidance on specific online safety concerns

**Online Grooming & Exploitation**  
Indicators of potential contact by adults with harmful intent, inappropriate communication, or exploitation risks.

**Warning Indicators**

- Receiving gifts or special attention from unknown adults online
- Being asked to keep secrets or inappropriate secrets
- Requests for personal information or images
- Pressure to meet in person or move conversations to private platforms
- Exposure to age-inappropriate content or sexual material

**Recommended Interventions**

- Immediate communication with student about online contacts
- Review privacy settings and conversations with student
- Discuss about warning signs and reporting behaviour
- Consider referral to designated safeguarding lead
- Involve parents/carers in safety planning
- Report concerns to appropriate authorities if threshold met

7

**Online Safety Guidance New Concepts New Issues**

**What to 'Take Away'**

School Internet Access is **Safe**

When Students are Asleep, they are **Safe**

**If Children & Young Adults have Unsupervised Access 'AT HOME' They may be 'Unsafe' SO**

*New Concepts will continue to Evolve*

*Virtual Reality has an important part to play in Education*

*Artificial Intelligence has an even more important part to play in Education*

*Nearly 'ALL' Online Dangers occur as a result of Social Internet Usage*

*Cyber Flashing is 'Real' – Offenders Cannot be ID'd but we can help possible Victims*

*Girls remain possible Victims of Sextortion 'but' Boys are now target by Catfishing & AI*

**Legislation May help but Education is SO Important**

**AWARENESS KNOWLEDGE ADVICE INVOLVEMENT**

- Screen Time
- Online Challenges
- Best Practice
- AI & VR
- Prejudicial Behaviour
- Self Perpetual Radicalisation



Thank You



**Jonathan Taylor** MSc

[www.onlinesafety4schools.co.uk](http://www.onlinesafety4schools.co.uk)

[onlinesafetv4schools@gmail.com](mailto:onlinesafetv4schools@gmail.com)