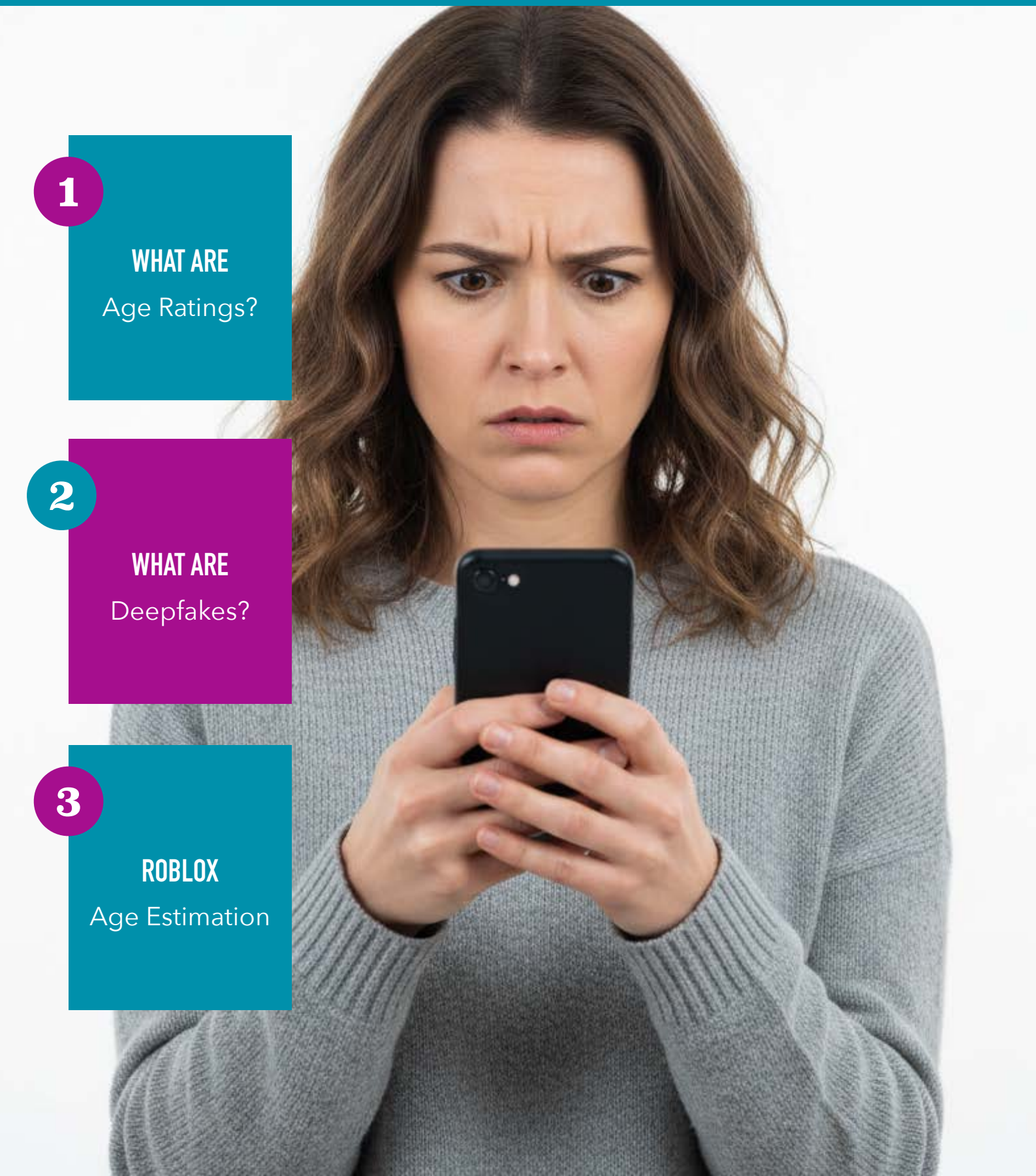
SAFE and SMART

Keeping Children Safe Online

October 2025



1 **WHAT ARE** Age Ratings?

2 **WHAT ARE** Deepfakes?

3 **ROBLOX** Age Estimation

What are:

Age Ratings

(and can you trust them?)

You would think that something as simple as age ratings wouldn't need an explanation, but everything is not as it seems. For a start, different stores might have different age ratings, an app or game rating on one store might be different in another store, and a really important one - does the age rating always reflect the actual requirements of the developer or company?

What are the common age ratings?

In the UK and much of Europe, most games are rated by PEGI (Pan-European Game Information). You'll see badges like PEGI 3, 7, 12, 16, 18 with small content icons (e.g., violence, fear, bad language, gambling, in-game purchases). PEGI shows the suitability of content—not how difficult a game is.

In North America, the ESRB uses categories such as Everyone (E), Everyone 10+, Teen (T), Mature 17+, Adults Only.

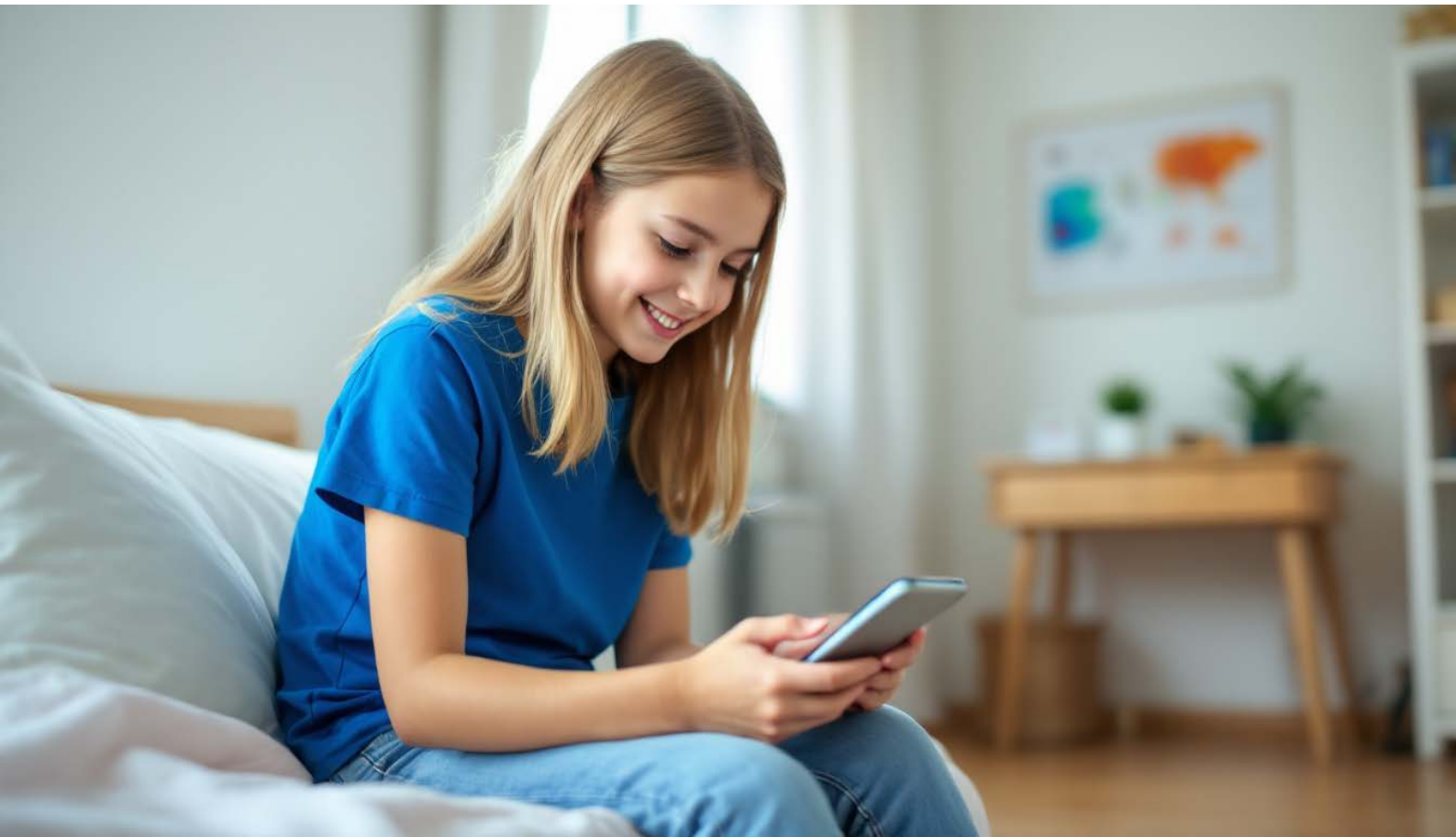
ESRB also adds “Interactive Elements” labels like Users Interact, Shares Location, and In-Game Purchases, which flag features such as online chat, location sharing, and spending.

On mobile app stores, the labels differ:

- Apple uses its own 4+, 9+, 12+, 17+ system;
- Google Play shows regional ratings via the International Age Rating Coalition (IARC) system (so in the UK you’ll usually see PEGI).

How are they useful for parents?

They’re useful, but not a solution. They’re an easy way to show the type of content your child might encounter. The badges and content/interactive descriptors help you spot potential red flags (e.g., online chat or in-app purchases) at a glance. But all children are different and we shouldn’t just base our decisions on an age rating.



Beyond the age rating

- **Content descriptors & interactive features:**
    - Look for labels such as Users Interact, Shares Location, In-Game Purchases / Includes Random Items. These don't change the age number but they can change the experience—especially contact with others and spending so they're always worth a look.
- **Data privacy information:**
    - Apple: open the Privacy "Nutrition Labels" on the app page to see what data is collected or linked to your child.
    - Google Play: check the data safety section for developer-declared data collection/sharing. You might be surprised about how much data is being collected!

Can you trust age ratings?

For the most part yes, but you should use age ratings as a starting point to determine whether a game or app is suitable for your child.

The reason for this is that recent UK complaints from the 5Rights Foundation and the Good Law Project in June 2025 have raised concerns that some popular apps display child-friendly store age-ratings, while the terms or privacy policies of the developer/company set higher minimum ages or allow extensive data processing/targeted ads.

Not only that, different stores can display completely different age ratings. It's always worth checking on Common Sense Media (link below).

Before you download

- Scan the badge and labels: PEGI/ESRB plus icons like Users Interact, Shares Location, In-Game Purchases / Includes Random Items.
- Open the privacy section: Apple Privacy Labels or Google Play's data safety—what's collected, shared, or linked to your child?
- Look for spending hooks: in-app purchases, subscriptions, and random-reward mechanics (such as loot boxes). Consider turning off or restricting purchases.
- Beware contact risk: if the app allows chat, friend requests or user generated content, plan together how to block, report, and mute by learning the features within the game or app.
- Check for any reviews on Common Sense Media:
  - Games reviews are HERE.
  - App reviews are HERE.

What are:

Deepfakes?

Generative AI (GenAI) refers to online tools (websites or apps) that can create new content such as text, images, audio or video from a few text prompts.

The more popular tools (called models) that people might be aware of are ChatGPT, Google Gemini and Midjourney, but there are now thousands of these models.

Deepfakes are AI-generated or AI-manipulated images, video or audio that misrepresent someone—making it look or sound like a real person did or said something they didn't. A good example would be a video of a celebrity on social media who is endorsing or advertising a product, yet that celebrity has had nothing to do with the product, their face and voice has been used, often without their permission, so the ad or endorsement is likely a scam. So too with political figures, where faces and voices can be used to spread misinformation.

The fast-paced nature of AI advancements means that these fakes are increasingly difficult to detect; the faces, voices and mannerisms are incredibly realistic. Every image within this newsletter is AI generated (apart from the Roblox image and the age ratings images).

And then there are nude deepfakes. These are exactly what it says - where clothing is removed or a person's face is placed onto an explicit body. These deepfakes and nude deepfakes now require little skill or time to create and more often than not target woman/girls.

HOW ARE DEEPFAKES BEING USED?

Deepfakes can have positive, fun or creative uses, but increasingly they're being used for harmful purposes. Internet Matters' research highlights sexual deepfakes (especially targeting girls and women) and growing misuse for fraud, scams and disinformation as a growing concern.

WHY THIS MATTERS

Apart from the obvious there are huge concerns about the ease of use, the scale, the fact that someone may become a victim and not even know they're a victim of abuse.

To give some context:

- In a 2024 survey for Internet Matters, **13%** of UK teens reported some experience of nude deepfakes (seeing, sending/receiving, or using/knowing someone who used a "nudify" tool). That's roughly 4 pupils in an average class of 30.
- **55%** of teens said having a nude deepfake of them shared would be *worse* than a real image, stating loss of consent or control, fear others would believe it, and not knowing who did it.
- The Internet Watch Foundation found **11,108** AI-generated child sexual abuse images on a single dark-web forum in one month. Evidence that this is an increasing issue.
- These nudify tools are cheap (sometimes free), widely discoverable via search, often lack age checks, and market "3-click" results. Incredibly, at one point there were a number of nudify apps on the Apple and Google Play stores a few months ago!

THROUGH THE EYES OF A CHILD

Anyone can be subjected to deepfakes, particularly if you use social media (including YouTube). From the perspective of the child we use the risk categories known as the 4C's. These are content, contact, conduct and commerce:

Content (What they see)

- Exposure to sexual deepfakes—of peers or celebrities—can be disturbing, humiliating, and hard to disprove once shared. Amongst some of the concerns of children and young people is the fear that others will think the fakes are real.
- Wider harms include misleading political clips or scams and misinformation that "sound like" someone who is trusted.

Contact (Who can reach them)

- Deepfakes can be used for sextortion (blackmail): threats to share fabricated nudes unless money or more images are sent. Boys highlighted how this could be used for bullying and blackmail.
- Some offenders will contact children and pretend to be someone younger, often stealing profile images from somewhere else to reinforce that young age. With the developments in GenAI offenders can now create their own images, they don't have to steal them from elsewhere anymore.

Conduct (How they act)

- Some boys have used nudifying apps to target female classmates, sharing fakes in chats and on social media. Not only is this illegal it is a huge violation of the victim.

Commerce (Money & data)

- As well as the potential for scams, nudify sites run on low-cost subscriptions, driving scale; some parents report filtering tools miss them. Teens also mention blackmail attempts as mentioned above ("pay or we post").

WHAT DOES UK LAW SAY?

- Any sexualised image of a person under the age of 18 (real or AI-generated) is illegal to create, possess or share.
- When it comes to sexualised deepfakes of adults, the UK Government announced plans in January 2025 to criminalise the *creation* of sexually explicit deepfakes of adults without consent. This measure is included in the Crime and Policing Bill, which (at the time of writing this newsletter in October 2025) is progressing through Parliament.

HANDLING NUDE DEEPFAKES

- With the advancements in AI and models being able to create ultra photo-realistic imagery, spotting a deepfake is becoming more difficult. Usually the most obvious is skin that looks too smooth, or maybe looks a little plasticky. That can't be relied upon as there are many filters within apps that can produce the same effect.

- Discuss with your child how fakes can be weaponised (bullying, sextortion, misogynistic “nudify” culture) and agree what your child would do if they see or are targeted by a deepfake.
- If you find or are informed about a nude deepfake of a child (whether your child or not), write down any evidence, such as usernames, website addresses. You can screenshot this evidence but make sure the original image **IS NOT** in the screenshot. **NEVER** share
- **Report quickly:** Use platform tools and reporting routes (see useful resources below). Fakes (indecent images) of under 18’s are treated as child sexual abuse material (CSAM).

WHAT TO DO IF YOUR CHILD IS AFFECTED

1. **Remove & report (under-18s):** Use Report Remove to get sexual images of under-18s taken down—even if they’re AI-generated. Outside the UK use the Take It Down service.
2. **Police route:** Report online sexual abuse or sextortion to the police. In the UK your best reporting route is **CEOP** (NCA).
3. **Blocking & moderation:** Report to the platform if they have a reporting facility.

Roblox at a glance

Roblox is enormous. In its latest quarterly update (Q2 2025), Roblox reported 111.8 million daily active users, up 41% year-on-year, with 27.4 billion hours spent in just that quarter - wow!!

Here in the UK, Roblox is consistently listed by Ofcom among the key platforms where children spend significant time online alongside other popular platforms such as YouTube, TikTok and others—so it regularly features in UK online-safety work.

Why do children love Roblox?

Mainly because it isn't a single game, it's a platform with so much to explore. Children can jump between millions of mini-games ("experiences") made by creators around the world. Some children also like the fact they can play and chat with friends, and also because of the creativity - many children learn simple game design and even scripting.

Why should parents be careful?

Most children use Roblox positively, but as with any large social platform there are risks, for example:

1. **Contact risks (strangers & grooming).**
   Roblox has public spaces and chat features. Predatory adults can try to befriend children, move conversations off-platform, or push for private chats/voice. Recent US state lawsuits have specifically alleged Roblox isn't doing enough to protect children from predators; Roblox disputes this and points to new safeguards.

2. **Inappropriate content in user-made games.**
   Roblox has a lot of moderation in place, but no system is perfect, especially given the number of users. Now and then, children may encounter unsuitable themes, role-play, or language before moderation catches it. Roblox says it proactively monitors text chat,

prevents user-to-user image sharing, and filters chat by age to reduce exposure.

3. **Commercial pressures.**
   Experiences often include in-app purchases ("Robux"). Without spend limits, children can click-through and rack up costs quickly.

4. **Peer pressure & scams.**
   Watch for "free Robux" scams not only within Roblox but commonly on YouTube videos as well. There's also pressure to trade items, or being coaxed to other apps/sites.

Age Estimation

In September 2025 Roblox announced that the company will expand age estimation to all users who access on-platform communication features by the end of the year. Instead of relying only on the birthday someone types at sign-up, Roblox will combine:

- Facial age estimation,
- ID-based age verification, and
- Verified parental consent (where applicable).

The aim is to know users' ages more accurately so Roblox can limit communication between adults and minors unless they already know each other in real life, and to apply age-appropriate features and content.

Roblox says this builds on over 100 safety initiatives shipped since January 2025 (for example: age-based communication tools like Trusted Connections, an AI system called Roblox Sentinel to detect early signals of child endangerment, and continued improvements to voice/text filtering).

Roblox also highlights its existing defaults for under-13s (e.g., no private chat/voice by default), proactive text-chat monitoring, and the fact that users can't send each other images.

With all that said, as Roblox acknowledges, no system is foolproof and there are still risks. We need to keep a watchful eye on what our children are doing on the platform, who they are friends with or who they are talking to.

Roblox can be great fun and creatively enriching, but it's also a busy social space, so set it up together, lock down the settings, keep purchases under control, and stay curious about what your child is playing and who they're playing with. That combination is the best protection on any platform.

Common Apps

This is not an exhaustive list, but tends to be the more popular apps used by children and young people.

Age requirements are set within the terms and conditions of the app provider, don't be confused by ratings in the app stores which can be

App	Age	Comments
	13	**Discord** - is a voice, video and text chat app that's used by tens of millions of people aged 13+ to tap and hang out with communities or their friends. Parental settings can be found **HERE**.
	13	**Instagram** - is a photo and video sharing app where people can upload photos, videos and messages to share with others. Parental settings can be found **HERE**.
	13	**Snapchat** - is a very popular app that lets users swop pictures and videos (Snaps) with others which are meant to disappear after they are viewed. There is also a messaging feature. Parental settings can be found **HERE**.
	13	**TikTok** - is a social media app that allows users to create, watch and share short videos shot on mobile devices or webcams. Parental settings can be found **HERE**.
	13	**Twitch -** is where people come together to chat and interact live. Think YouTube, but it is live rather then pre-recorded. Parental settings can be found **HERE**.
	13	**WhatsApp -** is a messaging app which uses text, images, video and voice record features to connect with others. Parental settings can be found **HERE**
reddit	18	**Reddit** - is a network of communities (called subreddits) where people can share information, their interests and hobbies. Reddit is an 18+ app, there are no parental controls.