

Preemptive Elicitation Model ('PEM') proposal

A remedy for societal divide?

Full research paper for model and condensed content listed on Love Your Hippo
Model by Dr Kelly Foxhall-Ridgeway
Last updated: 7 March 2026

Version publish dates on Love Your Hippo displayed on last page
The content within this paper is not to be used or copied in any way without absolute
permissions from the author.

Table of contents

References	1
Introduction	2
Harmful content	2
Research and evidence	6
Studies	6
Freedom of expression	9
Examples of negativity online	9
The Preemptive Elicitation model	12
Formula	12
How long is the space?	14
Other considerations	14
The disadvantages	15
Test cases	16
Conclusion	16

References

- 1) The reference '**Platforms**' collectively refers to the following applications through any device, but not limited to, Facebook (Meta), YouTube, Instagram (Meta), TikTok, Snapchat, X (formerly Twitter), Discord, Reddit, and LinkedIn, Roblox, Copilot, ChatGPT, Claude and Gemini and more.

Messaging platforms such as WhatsApp, Telegram and WeChat can be more isolated in the sense that the comments are limited to groups, and yield less traffic compared to a public post on social media. However, these platforms are not without their level of negative content and harassment and other incidents. There is still scope to share any form of content.

- 2) **MSM outlets** reference mainstream media outlets such as traditional media (newspapers, TV, radio), with websites and apps being the primary distribution method for a vast majority of journalists. This also includes the rise of new media as well as established brands (i.e. BBC, CNN, Sky News and so on).

- 3) **Generative AI** is a type of artificial intelligence (depending on the application used, also known as a Learning Language Model 'LLM'), that creates new content such as text, images, audio, code, or video by learning patterns from large amounts of data. Instead of retrieving information, it produces original outputs that resemble the examples it was trained on. It is conversational, but also can learn particular styles the more it is used. AI is integrated in many platforms, but it can also be a standalone application such as (as mentioned above) Copilot, ChatGPT, Google Gemini and so on.
- 4) **Sentiment analysis tools** use AI and Natural Language Processing ('NLP') to detect emotions (positive, negative, neutral) in text, such as customer reviews, support tickets, and social media posts.

Introduction

This paper details a proposed simple model which is intended to reduce the harmful content and negativity in triggersome responsive and reactive online content, **without removing the right for opinion and expression**.

The internet and social media are poster figures for modern times and while both prove to be exceptionally useful in millions, if not billions, of day-to-day lives, it is also responsible for being a conduit of behavioral change often leading to mental health decline.

While it would be economically detrimental to remove the internet for so many establishments, organisations and sole traders, it is now becoming very apparent that provisions, rules, processes and policies all need to be put in place to keep order, without overstressing our already overloaded law enforcement.

It is important for the lawmakers and the enforcers to provide a blueprint to help safeguard and guide society, as well as the future generations, but it cannot just be up to them to regulate the online world. It is also up to others to unite in a combined effort.

Harmful content

Harmful online content includes a broad range of user-generated material, images, videos, and text that can cause physical, psychological, or emotional harm to users, particularly children and young people.

Under the UK Online Safety Act, this includes both illegal content that platforms must remove and legal-but-harmful material that requires restrictions to protect users.

1. Illegal Content and Activity

Platforms have a legal duty to remove content that constitutes a criminal offence, including:

- **Child Sexual Abuse Material (CSAM)**: Including images, videos, and AI generated, animated, or drawn material.
- **Terrorism**: Content that encourages terrorism, promotes extremist ideologies, or provides instructions for violent acts.
- **Hate Speech/Crime**: Material inciting hatred or violence based on race, religion, sex, disability, or sexual orientation.
- **Incitement to Violence**: Content encouraging serious injury or violence.

- **Harassment and Threats:** Cyberbullying, stalking, and threatening communications.
- **Intimate Image Abuse:** Known as 'revenge porn', this involves sharing private sexual images without consent.
- **Cyberflashing:** Sending sexually explicit images (e.g. intimate areas) without consent.
- **Fraud and Financial Scams:** Phishing, identity theft, and fraudulent investment schemes.
- **Sale of Illegal Goods:** Drugs, weapons, or people smuggling.

2. Other harmful content

Several of these have just become illegal under the new legislation in the UK.

- **Self-Harm and Suicide:** Material that encourages, promotes, or provides instructions for self-harm, suicide, or eating disorders.
- **Bullying:** Persistent, targeted harassment that causes significant distress.
- **Pornography:** Sexually explicit material, especially if accessible by minors.
- **Dangerous Challenges:** Content encouraging stunts or virtual activities that are highly likely to cause serious injury or death (e.g., in apps or online games).
- **Misinformation and Disinformation:** False information, especially that which threatens public health or security, or undermines a democratic right.
- **Violent or Graphic Content:** Content depicting extreme violence, animal cruelty, or gore.
- **Sexism and Harassment:** Content specifically targeting a gender or non-binary.
- **Online Grooming:** The process of building an emotional connection to manipulate a child or vulnerable person for sexual or criminal purposes. Age of consent and capacity is involved with the determination of this.

3. Contextual and other behavioural harms

Harm can also arise from how content is delivered, or the interactions it promotes:

- **Algorithms:** Recommender systems that push users toward "toxic" content (e.g., in a rabbit hole of harmful content).
- **Addictive Design:** Features designed to keep users, particularly children, engaged for long periods.

The proposed model in this paper considers the following:

1. **Assistance/sentiment tools** to be added to platforms. References to these tools have already been included in several research papers as potential remedies, and could be implemented at the platform provider's discretion. If there is a chance of reduction of traffic then it is unlikely that tech companies will implement these unless forced to do so under amended regulations.
2. **MSM outlets** also need to be accountable for their part in the publication of misinformation and at times disinformation and malinformation.
3. The main focus of this paper is us, the user. **The responsibility is also on the users who are also the individuals contributing to these platforms.** We must accept responsibility for our actions (and that of our dependents if applicable). It is not just up to the technology companies\platform providers to keep adding more elements, restrictions and assistance tools.

While some of the technological advancements above would be effective, such as the sentimental analysis tools, it places the burden completely on the technology platforms to implement. Similar technology has been in place in, for example, litigation platforms for years and they do prove to be effective when searching for and tagging information, but the tools are, in a manner, telling us how to feel, what to do and sometimes influence an act that follows.

The platforms have a responsibility to provide a legally safe environment, but it is up to us to work on the ethical side and choose what we post. Where points 1 and 2 are omitted, point 3 regarding user responsibility and the proposed model, would be sufficient by itself in many scenarios. Whether we agree with it or not, there is already some regulation on these platforms.

The platforms are removing harmful/negative content, even if we agree that what they are doing is not enough, and by us sharing the responsibility is not necessarily 'victim blaming', because those that contribute to the platforms are already involved.

In their Q3 2025 report, Meta stated that it removes [90%](#) of harmful content.

'Of the hundreds of billions of pieces of content produced on Facebook and Instagram in Q3 globally, less than 1% was removed for violating our policies and less than 0.1% was removed incorrectly.'

'For the content that was removed, we measured our enforcement precision – that is, the percentage of correct removals out of all removals – to be more than 90% on Facebook and more than 87% on Instagram. That means about 1 out of every 10 pieces of content removed, and less than one out of every 1,000 pieces of content produced overall, was removed in error.'

In a news [article](#) on LBC, a spokesperson for TikTok stated that they remove most of the content *'Using robust detection systems and dedicated enforcement teams to proactively identify and remove this content, we remove 99% that's found to break these rules before it is reported to us.'*

'As a company, we comply with the UK's strict data protection laws.'

The argument of control and dictation comes into this heavily as where platforms are already controlling what a user sees through algorithms, by tasking the companies with further responsibility, it keeps them controlling the user. **If the user takes responsibility then they are given the opportunity to make their own choices rather than the choices being made for them.** It then becomes an equal relationship.

To continuously be controlled by the platforms could be seen as inadvertently discrediting the user by not having faith in them to make choices in their own lives, when many individuals have the capacity (legal sense) to make their own choices.

For the population that cannot interpret online social cues, behaviour can be learned over time, therefore if users begin making wholesome choices others are likely to follow. There will always be those who choose to be awkward and deliberately post harmful/negative content within the law (aka the 'troll'), for fun, or in rebellion to the norm.

There will also be those who post inaccurate content through the best intentions, but if they are the minority their online voice will soon be diluted and in time, they could even learn more constructive communication. By others not reacting, these individuals will likely tire (this is historic human behaviour). It may be that an algorithm will keep harmful content away due to it learning from the user that the user does not wish to see it. This does not solve the problem, it just moves it out of sight.

The topics of the law and online safety for children will not be exhaustively detailed in this paper, as the main focus is on the wider general usage of online platforms, and the impact the increase in negative content is having on our society.

It is clear that most of society has a new age **addiction** to technology and mobile devices, and its impact on society's behaviour, along with the world, is more than significant. The intention of these platforms was to drive engagement, connection and bring people together. However, the result has turned into a type of online content infestation, which is detrimental to society and is far more contagious than other **behavioral addictions** (such as shopping or exercising which are considered to be socially acceptable).

Technology can also be described as socially acceptable, and it is beneficial in many ways but, it brings with it a more dangerous behaviour addiction because it is unconscious¹. These platforms are designed to keep the user's attention with the dreaded infinite (or doom) scroll. The idea is to keep the user logged on, to keep their attention for as long as possible, but at the same time it is also invading one's subconscious² and reducing our focus and attention spans in other settings (e.g. work).³

Modern technology advances extremely fast and new updates are continuous. While new laws and regulations are being introduced to assist with online safety, these do not accommodate the addiction that is suffered. There are profits in this addiction to the platform providers and content creators and that is what keeps them and the users online.

What is not clear yet is if society is also suffering from some sort of **trauma**. Are people trying to escape something? An addicted society is not ideal, but combined with a potential trauma, could have an inestimably lasting effect on a portion of society. It is unclear how big that portion is, but there is still scope to change before it is too late.

Humans are more attracted to negative news, this is historic behaviour where the gravitation is towards gossip. However, recent times have shown a vast disregard for welfare (for others and self), and the neglect of responsibility, and/or accountability. Negativity is no longer contained, neither is diplomacy or subtlety it seems, the filter is off and it is all around us. It is environmental, and the environment or our surroundings has always had the biggest impact on the psychology of humanity (as well as animals).

The model described in this paper is to provide a way for individuals to take back **consciousness**, their **control** and **choice** with just a few minor changes as to how they conduct themselves online.

So what happens if we do not stop? What will happen to society if there is a prolonged exposure to negative and harmful online content? We can expect a mental health crisis (research examples included in this paper), normalisation of distress, heightened risk of aggression and antisocial behaviour, erosion of social cohesion and lack of trust with fact finding, desensitisation to harm, increased vigilantism (including the incorrect targeting of individuals), further distrust in publications, information and journalism, developmental issues in our youth and in extreme cases it could even lead to war (to a certain degree). The recent '[school wars](#)' is a current example of an organised online trend.

Unsettling isn't it.

¹ Addiction often originates as a conscious choice (curiosity, stress coping) but transforms into an unconscious, compulsive behavior driven by brain changes, such as altered reward systems.

² The subconscious links substance use with rapid relief.

³ Negative Distractions Most Disrupt Attention and Linger in Memory, 18 April 2025, [Neuroscience News](#)

Research and evidence

Some of the following articles contain distressing themes and rude language.

It seems somewhat humorous to publish a paper online when we are highlighting the negative impacts of online content, but this is not a bid to take away the internet or social media, it is to encourage people to take back their wellbeing.

There is substantial (both observational and statistical) evidence, and a wide body of research demonstrating the negative impacts of social media platforms and other internet sites. Studies show significant correlations between exposure to negative content, cyberbullying, and various mental health issues, particularly among adolescents and young adults.

Studies

A 2023 Norwegian [study](#) published by the Folkehelseinstituttet - FHI, found that 74.5% of secondary school students reported at least one negative experience on social media. These experiences include receiving hurtful messages, exclusion, or name-calling, which are linked to increased risks of anxiety and depression.

A survey carried out by the [Pew Research Center](#) in 2022 revealed that 46% of US teens experienced at least one form of cyberbullying behaviour, with 32% reporting to have been called an offensive name.

Stanford University published an [article](#) including survey results on 25 October 2024 detailing 'How Negative News Takes Over Your Feed', but also provided some solutions.

The lead author Brian Knutson said '*a sort of sentiment analysis tools could be implemented by users to filter out strongly negative and misleading content from their own feeds, but that puts the onus on the user and cannot prevent willful consumption of bad information.*'

This somewhat compliments the other side of this paper where we question why can't users take responsibility. Some platforms may not allow the integration of tools by users even if the users are trying to help themselves. The platforms are to provide a safe environment and modify harmful content on their servers, but people, as in general users, actually post the majority of the content in the first place. Therefore, is it not time we take control ourselves and mediate this ourselves even if this means in a more pedestrian way?

'The team also found that the most biased news sources, left or right, had roughly 12% more high-arousal negative content than balanced news sources and that these highly arousing negative posts were most likely to go viral.'

'Knutson concluded. "Perhaps by filtering content sentiment, along with the semantics and source, we can provide users with a new and useful set of tools."'

An [article](#) from the National Library of Medicine, showed Negative online news articles are shared more to social media data from four US and UK news sites (95,282 articles), and two social media platforms (579,182,075 posts on Facebook and Twitter (now X), and shows social media users are 1.91 times more likely to share links to negative news articles over positive.

The article mentioned that the impact of negativity varies by platform and for political articles, is moderated by topic focus, with users showing a greater inclination to share negative articles referring

to opposing political groups. Additionally, negativity amplifies news dissemination on social media to a greater extent when accounting for the re-sharing of user posts containing article links. These findings suggest a higher prevalence of negatively toned articles on Facebook and X compared to MSM outlets.

In 2025, an [article](#) was published detailing the experimental online study on the impact of negative social media comments on anxiety and mood. The study involved 128 adult participants (85 female, 43 male), and individuals were asked to share blog posts on a simulated internet forum.

Subsequently, they were exposed to either negative, neutral, or positive comments, and their mood and anxiety levels were measured using validated scales.

The results showed that negative comments significantly increased anxiety and decreased mood compared to neutral or positive comments, while gender did not show any significant effects. Younger adults reported stronger anxiety responses to negative comments than older adults, suggesting heightened sensitivity among younger users.

While this result may be unsurprising, social media algorithms have been found to promote dangerous content, including self-harm, promote eating disorder material, along with risky viral challenges, exposing impressionable young people to harm.

Ellen Roome is one of several bereaved parents campaigning in the UK for her son Julian “Jools” Sweeney, 14, who died in unclear circumstances in 2022. It is suspected that he died from a viral online challenge that went wrong. She has campaigned since his death and has been [successful](#) in demanding a change in the law to automatically preserve a child's data after their death (aka ‘Jools law’), as well as promote online safety for children in general.

Recently Ellen has been included in discussions with the UK government regarding the possibility of following Australia's implemented social media ban on 10 December 2025, prohibiting users under 16 years old from accessing major platforms like TikTok, Instagram, Facebook, X, and YouTube in Australia. The law places the onus on technology companies to enforce age verification or face fines up to millions of dollars.

Research from the [Youth Endowment Fund](#) (YEF) in 2024 showed that 70% of teenagers in England and Wales had encountered real-life violent content online in the space of a year, with many seeing physical fights or content featuring weapons.

The National Library of Medicine also published an [article](#) called ‘Won't We Know If It's Toxic? A Review of the Impacts of Social Media on Public Health’ stating:

‘With technological advancement and the rapid evolution of the internet, the target demographic of social media has changed drastically. Consider the lifespan of Facebook (now, Meta), which was originally only for college students (18 years and above). Now, 95% of US teenagers are using social media, and nearly half of these teens report being online “almost constantly” (doubling from 2015). In children 8–12 years of age, 40% report using a social media platform, despite 13 years being the minimum age for social media use, per the Children’s Online Privacy Protection Act (COPPA). Ultimately, social media continues to dominate more and more of our attention, with the average individual spending 151 minutes per day on social media in 2022.’

The Royal Society for Public Health (UK) published an [article](#) mentioning in its key findings that 91% of 16-24 year olds stated they used the internet for social networking. This article also states that

social media use is linked with increased rates of anxiety, depression and poor sleep, and 70% said they had experienced cyberbullying.

Their recommendations included an introduction of a pop-up heavy usage warning on social media platforms, to highlight when photos of people have been digitally manipulated, safe social media use to be taught during PSHE education in schools, social media platforms to identify signs of mental health problems via posts, and discreetly signposted to support, and professionals who engage with young people to receive social media training.

Some of these items are already being implemented in schools and the general population where schools are also sending parents and carers newsletters via email. This again places onerous on everyone else other than the user. In the scenario where the user is underage, parents and carers need to be involved for guidance with the monitoring levels being subjective to the child/adolescent.

Where the current method of schools sending multiple notifications to parents and carers, this too needs to be conditioned as too many emails will remain unread either out of forgetfulness or ignorance. We are in a 'too much information' society, therefore if a message is intended it is not ideal to send an email along with five others in the same day.

I attended a [lecture](#) held at The Royal Society by Professor Iryna Gurevych called 'How to spot and debunk misleading content'. Professor Gurevych was also awarded the Milner Award 2025 for her major contributions to natural language processing and artificial intelligence, which combine deep understanding of human language and cognitive faculty with the latest paradigms in machine learning.

Professor Gurevych included examples of misleading content which highlights false claims on social media supported by the misuse of credible scientific publications, images or videos taken out of their original context and paired with false narratives, and misleading charts designed to persuade audiences to accept inaccurate statements. Clickbait was also a prominent topic of this lecture.

The lecture detailed how we can identify and debunk misleading claims, and helped educate the audience on the tactics used to create deceptive content and demonstrate how the latest advances in machine learning and artificial intelligence, can be applied to protect both people and machines from misinformation.

This is a prevalent topic as a lot of organisations (including the legal sector which I work in), has been implementing generative AI into its general working day, but with strict controls to ensure acceptable use is in place. This combined with the seriousness of checking the content AI produces, as it can hallucinate, and taking responsibility for the distribution and its usage.

'74 suicide warnings and 243 mentions of hanging: What ChatGPT said to a suicidal teen', is an article by [the Washington Post](#), showing the posts between 16 year old Adam Raine and his ChatGPT account, where attorneys for his parents showed how the chatbot became a confidant as he planned to end his life.

Adam conversed on average five hours per day with the chatbot, during which ChatGPT used words like 'suicide' or 'hanging' as many as 20 times more often than Adam did each day. The analysis was included in a lawsuit his parents have filed against ChatGPT's maker OpenAI.

Chatbots and generative AI are designed to give an answer when you ask it a question, and it is only as intelligent as the information or training set it gains its information from. In this instance, the chatbot encouraged Adam to call 988, a national suicide lifeline, 74 times over a 4 month period, but it did not stop the conversation completely. If the technology is not trained to alert authorities, then it can only

work within its means. If not already in progress, this would undoubtedly open doors to ensuring that stricter protocols are added to the Chatbots conversation.

Columbia University Irving Medical Center published an [article](#) where an [investigation by the Wall Street Journal](#) revealed that Facebook was aware of mental health risks linked to the use of its Instagram app but also kept those findings secret.

It was found by internal research that Instagram worsened body image issues for 1 in 3 teenage girls, and all teenage users of the app linked it to experiences of anxiety and depression. Watchdog groups have identified [Facebook and Instagram as avenues for cyberbullying](#), and reports have linked TikTok to dangerous and antisocial behavior, including a recent spate of [school vandalism](#).

Overall, we can all publish an agenda online and even if there are good intentions behind it this can still lead to harmful content. We cannot deny, even from the handful of research studies above, that there is a societal crisis in place, however it is also clear that we want this to stop.

The University of Oxford published an [article](#) stating a new large-scale, global survey revealed that most people want harmful social media content such as physical threats and defamation to be restricted. This also applies in the USA where several social media platforms have recently modified their policies in favour of unrestricted free speech.

The data indicates that while social media can offer benefits like social support, the risks and prevalence of negativity on these platforms represent a significant **public health issue**.

These studies are a handful of examples over a small timescale that is recent, but it is clear looking at the results that as time goes on the percentage of people impacted by harmful and/or negative content continues to increase.

Freedom of expression

Freedom of expression is a fundamental human right allowing individuals to hold, seek, receive, and impart information and ideas without locality interference. It covers spoken word, writing, media, and artistic expression, serving as a pillar of democracy by enabling public debate and accountability.

The legislation will slightly differ depending on location. In the UK, the Human Rights Act 1998, [Article 10](#) covers the right to freedom of expression.

Unfortunately, this is often misquoted or misinterpreted on social media platforms as 'freedom of speech' because the description does differ depending on where the person posting is, but also the person posting/commenting will sometimes not check the validity of their text before posting. While others use the 'freedom of speech' concept as a 'free for all' to post whatever they like no matter how offensive it is.

Examples of negativity online

The below list some examples and excerpts of the comments/responses. Not all responses could be included due to the length and volume, but the posts shown below have a vast range of negative reactions/responses. No names have been disclosed in these examples for anonymity reasons, and in some cases the full comments could not be extracted due to inaccessibility of the group they were posted to.

16 January 2026, [LBC](#), 'Mother sues TikTok for answers over death of son, 14', posted on Facebook.

Tagline: 'A British mother whose son died in her arms after attempting an internet challenge is to take on social media giant TikTok in an American courtroom', posted on Facebook.

Replies:

'A British mother who likely didn't teach her child that eating laundry pods was dangerous and not to believe everything you see.... Is upset that her child did exactly that?'

'I've just tried to jump the fence on the Swiss border on my motorbike but failed. Who do I sue? Steve McQueen's estate, United Artists or maybe the BBC who recklessly broadcasted it?'

'SINCE HER SONS DEATH!!! it wasn't important until she realised HER failure as a parent!'

'Like me, watching the A-Team and then reenacting it! So didn't she teach him not to be stupid?'

'She isn't the first parent to have attempted to blame TikTok for their own failings, and she won't be the last...'

'You under 16 year old shouldn't have a TikTok account and if they have unfiltered access to the Internet you should be doing a better job of monitoring what they're doing.'

'Bit late now, she should have done something before hand'

The lawsuit in question is being carried out by Ellen Roome and fellow bereaved parents against TikTok because they have not been provided with data of their children's TikTok history. If Ellen and the other parents had this information, they may be able to understand what the children were watching at the times of their deaths and what led up to it. The above example shows how people responded without reading the article and how the article is promoted with an ambiguous triggering headline, designed to hold interest and provoke reaction and influence unhealthy debate.

While people are entitled to their opinions, it was clear that once someone started posting negative comments, it opened the doors to others.

24 March 2025, ITV Anglia article [posted on Facebook](#) by a London wildlife group, 'Four alpacas have been shot dead in an attack at a farm in Cambridgeshire.'

Replies:

'There's a borderline serial killer on the loose. People who hurt innocent animals just should be erased from this earth through death penalty measures. Absolutely horrific and unfortunately nothing good will come out of this human.'

'What with those f@ cking slingshots?'

All these reports should be sent straight to the newspapers for circulation.

The report should shame the police and the culprits and authorities for doing bugger all to stop it!

This is going to escalate into a pandemic of zombie minded copy cats with zero respect for life .

Then down the line we will be the ones paying for it!'

'phycopaths.. The dirty no good evil bastards, the fuckin courts NEED TO MAKE EXAMPLE, out of the evil vermen'

This is an example of how an emotive topic can start an online commotion. As there is far more traffic online viewing these comments, more people can contribute which can lead to a modern day witch hunt. There were other commenting on this post that mentioned the police had no substantial evidence or lines of inquiry and therefore couldn't proceed further, but some of the posters (who posted some of the replies above), did not want to acknowledge this and started accusing those who were rationally responding with facts as 'sad' and accused them of endorsing animal abuse.

Anything to do with young children or with wildlife substantially increases the emotional responses and some of these comments escalated to the point posters said those committing these acts needed their limbs removed⁴.

There is no shame in feeling sadness when an animal or human is hurt, it is how far a commenter goes that determines the difference between opinion and eliciting hate speech. It is not a case of one rule for some of society and another rule for others, we are all answerable to the law. Unfortunately [Lucy Connolly](#) found this out in 2024.

There are psychological reasons as to why someone's responses change in emotional situations, and how these are escalated behind a computer screen, but posting comments such as the ones above are not needed and plus there is no end result accomplishment.

In the Facebook group 'British Airways Complaints Advice group', a long post was displayed on 23 March 2025 stating that the UK airline British Airways did not take food allergies on flights seriously.

'Moan moan moan. Always one somewhere'

'Suggest you seriously reduce dramatically the volume of your post - you cannot seriously expect your audience to plough through such volumes of script. I fell asleep after the first hour . . .'

'Well thanks to the minority of people who don't like nuts, I can no longer get a bag of peanuts on a plane.'

'This post is so funny, they are flying from India, did they not eat anything in India? No food hygiene standards in the entire country.'

'I sympathise with allergies. How ever they are your problem. I object to being told I can't have nuts with my drink, where does it end. In a restaurant they would not ask clients to not eat the nibbles. ! Carry your epi pen.'

Allergies have gained more awareness in recent years, but they took a long time to integrate into society. Toxic allergies such as nuts are more known because of the documented risk, but there will be audible frowns when an announcement is shared on a flight, even though this can be a life threatening condition for many.

The lack of empathy online coincides with depression, anxiety and heightened stress which is again unnecessary.

⁴ Cognitive dissonance refers to the mental discomfort we experience when our beliefs, values, or behaviors are out of alignment. This uneasy feeling motivates us to resolve the inconsistency either by changing our actions or by rationalising and adjusting our thoughts. As a subtle yet powerful psychological force, cognitive dissonance shapes our choices and convictions, often operating beneath our conscious awareness.

The Preemptive Elicitation model

Based on already existing paradigms, the Preemptive Elicitation Model takes it further for modern, where preemptive questioning can avoid a reaction to triggering topics, which are enhancing the drivers of negativity resulting in societal divide from online content.

It sounds too easy, but if the template is hard then it will be unattractive and arduous leading to redundancy.

The model is something very small but could be quite powerful in the long run but it will likely take some time to form the habit which is an unavoidable manual process. Therefore emphasis may be needed on the rationale of this model. For example, if people know they are potentially suffering, which could result in long term health problems, then they may be more forthright to change.

This behaviour is likely to be hard to adapt to because scrolling online is now integrated in everyday life, but with a small change the results could be substantial.

With the focus on perception and self awareness, decision making and how our behaviour affects the present, as well as the future, this model reminds everyone that humans are capable of humbly dealing with rebuttal/responses, and are also capable of existing without rewards (further information below).

Formula

The following gives a present day example, but the model is not just limited to social media.

Responses to social media posts and/or comments

The post and/or comment has to be triggering, such as an emotive topic, something personal and/or something that leaves a lasting feeling (e.g. a clickbait topic). A bland, neutral non-attentive post does not incite the same reaction.

While the terms 'react' and 'respond' are widely used interchangeably, there's actually a significant psychological distinction between them.

Reacting is our knee-jerk, automatic response to a stimulus, often driven by our emotions and instincts. It's like hitting the 'fight or flight' button without much thought.

Responding, on the other hand, involves a more deliberate process. It requires emotional regulation and thoughtful consideration before taking action.

Once the trigger is in place (such as an emotive post or comment) a user can either:

1. Not respond; or
2. Respond (it does not matter if this is a delayed response).

When a Response happens, two more routes can take place:

1. Constructive comments from those who are either self-aware or who are not linked to the initial trigger directly, or a reaction such as a 'Like'. This is a **minimal response**.
2. Those that have a further triggering reaction because the topic has happened to them, or there is some other direct connection. This is a **major response**.

The more triggering the post is, the more it will be responded to, and/or shared ensuring that there is a risk of reactive negativity which then spreads, mutates and can be influential. If you have a campaign you believe in, or have a strong opinion about something, then a type of impulsive motivation to share the post quickly can happen. Sometimes this is from good intentions, or a high alert reaction to something, for example you must share to warn others, or the sharing is to discredit someone or something and gain more like-minded support, much like a pack mentality safety in numbers manoeuvre.

The danger being is that you do not need evidence to respond to a post, many do not rely on it 'in the moment', and the more triggering a post is, the stronger the traffic/footfall will be. The non-factual comments can also be spread masquerading as fact because the poster believes it, which promotes misinformation.

MSM outlets also use this method to gain more traffic to their posts as they know this strategy works which is profiting from peoples feelings.

If there is a reaction, the reactor could be missing complex parts of the original topic. They may lack self awareness, and not be interested in facts especially if they are angry. Responders will sometimes comment for humour purposes rather than be in agreement or disagreement.

Common process: **Trigger – React/Response – Realisation.**

The Preemptive Elicitation Model process: Trigger – **Question – Optional response**

The **space in between the Trigger and Response** has been acknowledged in other papers, but not utilised as well as it can be.

If the response or reaction is organic and immediate, then the motivation for investigation (for example questions) will be disregarded, when these further questions can help avoid unnecessary conflict and validate the original post.

In this situation the reaction time is not immediate, unlike laughing. Laughing is immediate if we hear something that is funny, it is a more instinctive reaction, whereas an impactful post or comment that triggers us, even when we respond quickly, there is still a small space of time which can be utilised e.g. this can be the time it takes to type a response and click send.

By asking ourselves one simple short question, it will hopefully delay our response.

The question one can ask in this space before responding:

Do I need to respond?

By adding the word '**respond**' rather than 'react', it gives a more positive feel from the onset, so we are programming ourselves subconsciously to try and be more positive rather than negative.

The word 'response' is described as a more thoughtful, intentional, and considerate action.

The question triggers a pause, to process emotions, with scope to initiate the 'thinking brain' and choose a mindful way to act, aligning with one's desired outcome.

A desired outcome may not be a positive one. It may be the intention to still be negative but the idea is to denormalise hasty responses/reactions in the first instance.

If after asking ourselves this question, which will inevitably make us pause for a certain amount of time, we can then engage the 'thinking brain' and we then may be able to move on to the next short question.

If an individual does intend to respond, the second question would be:

What might the outcome be?

This question will also allow us to rewire the tone that we deliver in our response, even if it is more of a reaction or it is more of a heated response, it gives us that time for the thinking brain to engage and then later provide a more catered thoughtful comment.

It may be that a quick response is required but at least at this point we have used that space available between the **Trigger** and **Response** to have the opportunity to think, and to choose what the response would be.

How long is the space?

So what is the small space in terms of timing?

According to [Very Well Mind](#), the 'thinking brain' (prefrontal cortex) can begin to step in and override the 'emotional brain' (amygdala) in as little as **6 seconds after a trigger**, though the initial emotional, physiological reaction is nearly instantaneous.

Other considerations

When we lose control over emotion we can enter a highly destructive cycle. The emotional area of our brain can override the rest of the mind unless we take control. This does not mean there is something wrong with us, we are human afterall, but this is more about engaging the 'thinking brain', or the thinking mind, over the emotional one when **interacting with people we don't know on the internet**.

To provide further context, the thinking mind deals with logic, speech, planning, impulse control, judgement and reasoning to name a few. The prefrontal cortex is the area of the brain that can help or hinder communication. The frontal area of the brain provides this over highly emotional reactions.

Often there is a need to respond in order to gain a **reward**, but the reward isn't really a necessity when we engage online, it is a want, a nice to have.

Humans will want rewards to sometimes feel validated, or to have the last word in this instance, but if less people declined the reward, then the spread of influential negativity would minimise, and the likelihood of societal divide would also not progress.

There is a severe want in recent times to document thoughts online, and to show off what we are doing or what we have.

Our online world is very much a glamour of our real work at times, or a showcase of our lives, and this is a contagious type of learned behaviour from others who are also online using the same platforms like we have conditioned ourselves into a new superficial competitive norm. It even extends to where

videos are being viewed that have been staged but appear to be real, and many of the audience believes it is reality.

When designing this model, the concept of using a tool to delay comments was thought of, but this also negates the process of organic thinking which is better for mind growth, rather than using third-party apps or approaching social media platforms to apply such a function as it may impede their mission statement.

However, in 2025 Facebook began testing a 'thumbs down' icon on random posts which supersedes a test 'dislike' button in 2017 in Facebook Messenger. The button is intended as a 'downvote' system to help algorithms identify and suppress spam or disrespectful and abusive contents. While this button is anonymous to other users, it has been criticised for inviting more negativity.

The disadvantages

The problem is how do we implement a trigger (excuse the pun) for these questions, other than relying on human habit forming.

Individuals have to remember these questions, but they are easier to remember than a set of coping strategies and once used a few times they are likely to remember again rather than using a repellent response or base the model on consequences.

We could suggest that the platforms have a way of prompting the question when you go to respond to a comment, but this again takes control away from the user, and gives the platforms the opportunity to sway technological control, it could also see a decline in traffic from people choosing not to comment at all.

This model also alleviates the dependency on generative AI, which is different from using it for what it is intended for and that is to assist you with specific tasks. Using these questions to prompt your response means the decision is purely organic so it is delivering a more constructive message which could be more positive in general and help promote a growth mindset.

As with all models and experiments there will be disadvantages and/or challenges. The challenge here is that you simply cannot catch them all (yes that was a Pokemon reference). The reason being is that we are all different for a reason and that's what makes us special. Humans have different behaviors, but commonly we can all instinctively respond and/or react to something that impacts us personally.

There will be the reluctance to change from some and while many people could benefit from mindfulness and self-awareness, as others will not take this because they feel they do not need it or to be awkward/repel.

Our brains were created for survival and still think that we are in the age of cavemen\neanderthals fighting for our food but we are not, we simply go down to the supermarket now to get our food. There is no fight so it's kind of like our brains are looking for that fight that it was programmed to originally do. It is not socially acceptable to start random fights walking around outside, so we do it online with many thinking they have a valid excuse because they are behind a screen.

It is unreasonable to suggest that this model is intending to take away everything that triggers us, this is not the case. We are entitled to have our opinions and have reactions, but we severely need to reduce the amount of things we are reacting to, because we have got to the extreme where the online world is spilling out into the offline world.

It also causes a fundamental distraction to the point where we are not as efficient in the workplace, as parents/carers, or as efficient humans in life in general because we have taken on this level of escapism.

Test cases

I have been using this model, and I used myself as a test candidate because I regularly need to have the last word. I was bullied at school and in my adolescent and adult years I always said that I would never allow myself to be downtrodden.

A lot of the time when deciding whether to respond to a post or not, I chose not to. I always had a fear that if I did not respond then those that were publishing misinformation (for whatever reason), would be seen as being correct, and often this misinformation was about a certain organisation and it wasn't true, so I was worried about the lack of faith people would have in them, and the vigilante effect, so I used to constructively reply, with no negativity, but then found myself in an unintentional argument because the other person refused to believe what I was saying.

My husband tried this model, but he is of the notion that he will say what he wants in response to people and they will have to accept that he has an opinion as do they. Due to my husband's job, where he can be authoritative, I am not surprised by his stance on this. More work is needed!

Conclusion

Human behaviour is not uniformed, it is vast and wide and I could probably spend thousands of hours writing many papers on this topic alone. With this in mind, to be successful with this model, the ideal method is to engage the technology platforms to tighten their infrastructure further, but also to give control to the users which it seems no one is considering. By continuously taking that control away, then the users will still be controlled and somewhat dictated to, which does not change or solve the current situation.

I expect that if this paper is debated and ends up on a public forum there will be comments which state that the model is 'old news', and it's something that has been done before but that's the beauty of it - it has been done before. We know it works. We as humans need to go back to basics. We need to be responsible for our own wellbeing, for our own conduct and our own actions, while other entities can take responsibility and do have a share in the responsibility, we need to use our initiative as the choice has been with us all along.

For those of us that are parents and carers. We always tell children to be careful online, they need to be mindful of who they communicate with, and we can even lock down the content, but we also need to guide them and let them take control of their responsibility when it is appropriate - this is no different than when we let them walk on their own somewhere for the first time.

It can be difficult as quite often we feel hypocritical, I know I play musical screens on a daily basis for work. But it is also telling our children why we do this and even if we are playing games it is setting an example to do this portionately. There is nothing wrong with adopting technology, it is very efficient, but it is where we draw the line.

Even if 100% of the dangerous content is taken down from these platforms, there is still the dark web. There are still other ways. Removing illegal content is easy, moderating ethical or unethical content is another thing altogether.

Internet users' intentions will differ from general engagement, to catching up with friends, pursuing information for research or studying, then there are those who go on social media to stir up trouble. They add laughing emojis to posts that are not funny, and these along with other comments made by humans and artificial bots.

Overall. When was the last time you felt happy going on social media and commenting without thinking; *am I going to get backlash from this?*

We are all guilty of reacting rather than responding and in some instances a reaction is certainly called for, but it is now overtaking society to the point there is a clear divide on everything. In terms of clickbait the increased spread of misinformation where people rely on headlines instead of reading articles, and in the same cases, the headlines are purposely ambiguous, relying on people not to find out information.

Misinformation should be much lower given the amount of information we have at our fingertips but we now have virtual witch hunts, cyberbullying, virtual harassment, death by virtual challenge, and these all outweigh the virtual unity we need to do something to resolve this. Due to there being too much information we then ignore it all and hope for the best, but we don't switch off instantly.

The examples given above have mainly focused on social media posts, and responses to comments, but you can use the model for general internet usage, product article reviews, anywhere where there is online engagement.

Our brains were trained to multitask in school. We went from lesson to lesson. We learnt several things at once and sometimes we would be learning the lesson content as well as doodling in our books, so why do we have a problem multitasking now? Scientists are telling us the brain cannot do this, it can only work on one thing at a time. Given how different sections in the brain work at the same time regularly, how can this be?

Could it be that our ability to multitask has reduced due to the use of smart devices and the internet?

There are studies that suggest we may have reduced our IQs by several points due to the use of mobile devices and continuous multi (device) tasking. With a reduced usage of these devices along with doom scrolling, our focus and attention could substantially increase, along with our manageability to multitask, and we would lean on a growth mindset instead of fixed.

If society is suffering an addiction, which it is very likely, along with a trauma then it will be a long road, but the good news is that we are all responsible for our behaviour and we can control our wellbeing. If it gets to a point where society is completely dependent on these devices and the internet, then we run the risk of declining our own self-awareness, our own responsibility and eventually our independence.

'The divide' is here more so than ever before and it is likely leading us to a mental health crisis which could have severe long term impacts.

A healthy brain in our senior years may ward off dementia or prolong a good quality of life with dementia. Healthy emotional intelligence helps with immunity making us generally healthier. When you are happier you can also accomplish a lot more. It is like exercising, you typically receive better results if you do something you enjoy.

For those that do not want to change their ways online, there could be an element of disbelief, they may find it comical, or they do not take the online division seriously or it could be that they already have control, which is fantastic.

The concept of the '**space**' in between the trigger and response is central to ideas found in stoicism, mindfulness, and emotional intelligence. The main point is that by becoming aware of this gap between a trigger and response, one can move from an impulsive, automatic reaction to a considered, conscious response, thereby exercising personal agency and control over one's behavior.

Techniques such as mindfulness meditation and journaling are suggested to help 'widen' this space but this is subjective to the individual.

The Preemptive Elicitation Model, allows people to take back their control, consciousness and choice when engaging online, without removing the right to opinion and expression.

"Between stimulus and response, there is a space. In that space is our power to choose our response. In our response lies our growth and our freedom." [Viktor Frankl](#)⁵

Dr Kelly Foxhall-Ridgeway

Dated: 7 March 2026

The content within this paper is not to be used or copied in any way without absolute permissions from the author.

Date publication/iteration history:

27 February 2026

3 March 2026

7 March 2026

⁵ Viktor Emil Frankl was an Austrian neurologist, psychiatrist, philosopher, and Holocaust survivor, established logotherapy, a form of psychotherapy that identifies the pursuit of life's meaning as the primary human motivation.