“Online Content or Online Emotion: What is the Real Threat?”
(Manchester University Centre for Digital Trust and Society, 2 July 2026)

Introduction

1. The title of this conference is so on topic it is almost scary. I approach my address with excitement and some trepidation.

2. The stuff we are going to talk about at this conference goes to the core of a functioning democracy.

3. I am particularly delighted to be speaking to you because – unlike so many people - you are putting the online domain front and centre.

4. There is so much to understand and discuss. But I fear we lack the language to do so. We still lack the language and concepts to address the online domain.

5. Change is going to come thick and fast. I remember saying on one radio show, to scoffs, that the Online Safety Act 2023 was probably the first of 15 acts of Parliament to deal with the online domain. I think I underestimated.

6. Right now, pressure groups, popular sentiment, newspapers and a group of ‘tech lords’ in the Upper House are changing the rules and the government and civil service appear flat-footed. Agree with under 16 ban or not? Light touch regulation for AI or ban chatbots? Fine the platform bosses or bet the UK’s growth on big tech?

7. New rules are coming down the track. We need to find a way, collectively, as a society, to make informed choices.

8. And this is a challenge to democracy as we know it. Laws are ordinarily formed by a slow process involving bills in Parliament with consultations and select committees and time to amend and push back. Fast laws, emergency laws, rule by secondary legislation, these are sometimes justified but they are also fraught with the risk of unintended consequences.

9. If we are going to make good choices at the pace we need to have an instinct for what is right and what is not. We need a value system. This is where academics and civil society and think tanks and above all politicians need to take a lead.

10. It is no longer sufficient to say something is problematic and further study is needed.

11. What I want to do in today's address is to suggest that emotional manipulation is a useful benchmark when considering the threats posed by the online domain, and that conversely disinformation is a concept of limited use if any use.

12. In short, if one's approach is about information and disinformation, then it becomes a debate about free speech where the absolutists have some good arguments and, currently, a great deal of the power.

Why threat?

13. I have two jobs. My first job, since 2019, is Independent Reviewer of Terrorism Legislation. There is no doubt in my mind that the online domain has changed the nature of the domestic terrorist threat. People form ideologies, feed grievances, consume propaganda, plot, encourage, glorify past attacks, and become obsessed by the vocabulary and imagery of death, and this leads them into terrorism.

14. I refer to the online domain as a whole, not just social media, or terrorist operated websites. All of it is relevant. Gaming is relevant for recruitment and reenactment of massacres. Old-fashioned bulletin boards feed extremist obsessions. YouTube allows terrorists charm the senses with extremist nasheeds.

15. My second job, since early 2024, is Independent Reviewer of State Threat Legislation. The simple point here is that foreign powers, and their proxies,

make use of the technologies of the day. An important current feature is the recruitment of saboteurs on Telegram.

16. So too is the exploitation of our inability to leave our devices, providing an opportunity to pipe content into our brains that advances the strategic objectives of foreign powers. Some of that exploitation is planned, but much I am afraid is simply self-inflicted.

Examples

17. I'm going to start small. This year, I met two young people who were serving long prison sentences after they plummeted down the Extreme Right Wing Terrorist rabbit hole as young teenagers and committed terrorism or what I will call terrorism-adjacent offences.

18. The first had already developed, entirely offline, an interest in a particular form of violence. With the availability of information online, that interest became an obsession. This led them to search for and find an online discussion forum dedicated to this form of violence. The forum was tinged with political extremism and calls to action. An interest in weaponry developed, and the police intervened.

19. The second was, if they can be believed, simply interested in politics. They were using a common social media platform, looking at a plain vanilla political story, when the algorithm started to serve up edgier and edgier content. This child DM'd some of the content creators, and within days they had been invited to an extremist website, and in due course this child who entered began to plan real world attacks.

20. At a group level, in 2024/5 the most prominent Al-Qaida media operation was Al Qaida in the Arabian Peninsula's Al-Malahem Media Foundation with its Inspire magazine[1]. Despite being based in the Arabian peninsula, AQAP propaganda focused on the Gaza and Israel conflict.

[1] Thirty-seventh report of the Analytical Support and Sanctions Monitoring Team submitted pursuant to resolution 2734 (2024) concerning ISIL (Da'esh), Al-Qaida and associated individuals and entities.

21. Another Islamist group, JNIM, rebranded its image to appeal to younger audiences. Importantly, to radicalise and recruit it exploited what the UN monitoring body described as “regional and social grievances”.

22. The exploitation of grievances is exactly what foreign powers do.

23. According to recent analysis by the Royal United Services Institute[2], whilst the Kremlin still produces original content, it spends an increasing amount of effort on amplifying existing inflammatory content from genuine users. This could be by using bots to repeatedly repost and engage with genuine content. It could be by using bots to generate provocative commentary on anger-making content.

24. Posting such commentary can give an algorithmic boost to the original content and the commentary – both onto the feeds of likely supporters and those who are likely to be strongly – and angrily – opposed to it.

25. The report identified postings relating to provocative terms associated with the protests and riots which bore telltale signs of Kremlin interference. Unless you’ve had your head in a puddle for the last 2 years you will recognise some or all of these terms: #twotierkier, #twotierpolicing, ‘UK has fallen’.

26. Russia linked profiles had increasingly turned to reposting Tommy Robinson and provocative anonymous accounts. In the words of the Report, “Most of the content was produced domestically; the Kremlin simply lent a helpful megaphone.”

27. To spell this out, Russia wants the UK to feel tired and broken and distrustful of our institutions and our traditional allies; China wants to denigrate Western democracy as a political failure; Iran wants to sew threat and fear against its enemies in the West.

[2] Morley-Davies, J., ‘How Did Foreign Actors Exploit the Recent Riots in the UK?’ (RUSI, 28.8.24).

28. The point from all these examples is that what really counted was not information or disinformation but emotion.

29. Take the 2 children. For both these isolated children the emotional aspect of belonging and identity was key.

30. To quote the recently retired director of Europol, “We have a young generation which is slightly detached from their parents, that are educated online, often by social media or gaming platforms…A young person who is still developing an ethical and moral compass, who hasn’t found their place in society, is psychologically more vulnerable to approaches by somebody who gives them attention, who gives them care, who engineers a way into their life and gains their trust”[3].

31. For AQAP and JNIM emotion is generated by exploiting local grievances. They are not providing information or disinformation so much as cottoning on to heightened emotions and a sense of unfairness and then offering an extreme solution.

32. For the Kremlin, it is the same. The material they generate or increasingly just promote is all a shortcut to emotion – whether the emotion is in support of Tommy Robinson, or against. What counts, and what allows the Kremlin to go to work like this, is emotion.

33. Best of all, as the philosopher Eric Hoffer wrote in the 1950s, is hatred[4]. The unifying effect of hatred is key to those mass movements that undermine our security.

Emotional manipulation

34. Too often policy makers and academics analyse online content as if they are critiquing articles in a journal. There is a reason we reach for phones to 'relax' after a hard day. It is not rational, it is emotional.

[3] Warrell, H., 'The teenagers enlisted as agents of mayhem by Russia and Iran' (Financial Times, 5.6.26)
[4] ‘The True Believer’ (Harper & Row, 1951).

35. We know this but perhaps have not fully absorbed it.

36. When tech companies want to keep us online or change our engagement in service of advertising revenue they don't go philosophers or sportswriters or others purveyors of valuable information. They go to psychologists.

37. There is a whole field of persuasive technology involving pings, double ticks, likes, infinite scrolls, addictive features, dark patterns, very few of which are about the accuracy of information but all about the manipulation of emotion.

38. I accept of course that riding on this emotional underpinning is content which conveys information. Some of this information will be true, some will be false, some will be partially false and partially true, but what determines whether it reaches people who might be motivated to carry out an attack, or lose doubt in democracy, is emotion.

39. Rage-bait gets clicks because it fires up our emotions. Conspiracy theories get clicks because they fire up our emotions.

40. It is well established that content that evokes high-arousal positive (awe) or negative (anger or anxiety) emotions is more viral. Content that evokes low-arousal, or deactivating, emotions (e.g., sadness) is less viral[5].

41. Indeed, it can be argued that the true measure of the impact of the online domain is reach not speech, where reach is based on emotion not truth or falsehood. I will return to this.

42. In January this year the journalist Fraser Nelson considered the algorithm that powers X or Twitter. This possibility arose because Elon Musk, to his credit, had done what no other social media company had done, and published his algorithm.

[5] Berger, Jonah & Milkman, Katherine (2012). What makes online content viral? Journal of Marketing Research 49 (2):192-205.

43. As Fraser Nelson wrote, this algorithm does not and cannot judge if you regard a post as interesting, insightful or useful; true or false. All it can do is gauge reaction. To “reply” is normally a sign of disagreement, and that’s rated as 27 times more valuable than a “like”. If a post leads to a debate: that is 150 times more valuable. This creates a bias towards content that, in his words, ‘riles’.

44. A different study concluded that Twitter’s engagement-based ranking algorithm amplifies emotionally charged and hostile content that users say makes them feel worse about their political out-group[6].

45. Of particular interest was the finding that the algorithm operated contrary to users’ stated preferences. In other words, algorithms rile us to keep us online, even if we would prefer not to be riled.

46. Unfortunately, there are plenty of entities – whether political or commercial or foreign power – who know how to exploit this algorithm and thereby exploit our emotions. It is sometimes as if manipulative content creators and the platforms we use are in league against us. I use X, which I like, but I do feel it is laying these landmines all the time. So many tweets starting, “What happened next will outrage you!”.

47. Conversely, if naively like me you post informative tweets which do not rely on emotional salience, you usually sink to the bottom. Some of my wisest pronouncements, it is true, have been seen by fewer than a hundred people.

48. The key point is that although I am responsible for my speech, I am not responsible for its reach. That is decided at an emotional level. It is a function of emotion not accuracy. Because my Tweets are careful and balanced and based on evidence, they have, I fear, much less cut through that if they were partial and emotional and manipulative.

---

[6] Smitha Milli, Micah Carroll, Yike Wang, Sashrika Pandey, Sebastian Zhao, Anca D Dragan, Engagement, user satisfaction, and the amplification of divisive content on social media, PNAS Nexus, Volume 4, Issue 3, March 2025.

49. It is important to add that this is not just about social media algorithms. The hook of anger has been exploited for years in non-algorithmic sites like bulletin boards or games platforms or discussion groups.

50. In short, because of the role of emotion, the whole online domain is ripe for exploitation by terrorists and hostile states who spot local grievances and promote salient conspiracy theories, creating emotional engagement with their worldviews for tactical and strategic ends.

Disinformation and Misinformation

51. I turn then to a very common mode of viewing the threat, as disinformation.

52. For reasons that are not immediately clear, these are terms that are often associated with the political left. The joke on the right is that disinformation was invented by liberals to explain President Trump's first election victory. Liberals just couldn't believe people would vote for him, unless they had been misled, ideally by Russians or some other outside force, because that preserves the illusion that people in democracies are otherwise good.

53. Politics aside, there are substantial reasons to be sceptical about the term disinformation. I am not saying there is no such thing as objective truth. There is clearly false information, such as the false information that the Southport killer was a Muslim called Ali Al Shakati, a false allegation which was carried on a wave of anger, ignorance and anti-Muslim hatred onto thousands of feeds.

54. But information is very often neither wholly true or wholly false, still less 'good' or 'bad'. It is very difficult to practice what I have seen described as "information hygiene"[7] because in most cases there is no acceptable mechanism in a free country to determine what is good and what is bad.

[7] 'Coronavirus: Here's how you can stop bad information from going viral': BBC Trending (20.4.20): "Meanwhile, experts are calling on the public to practise 'information hygiene'."

55. It is worth reiterating that during Covid the possibility of a lab leak was derided as false information, but has subsequently emerged as a credible explanation, indeed the most likely explanation according to the CIA[8].

56. To be fair, opponents of the lab leak theory may have been motivated by the need for global solidarity in the face of a pandemic, but such a motive merely illustrates the point. It is always tempting to the authorities to want to control information because of the impact that even true information can have on, say, public order.

57. Then there are other so-called conspiracy theories – think about all those conspiracy theories about the role of US intelligence in Iran in the 1953 coup – that later turned out to be true.

58. This is before you get the point that the fact of disinformation – the fact that someone has told a lie – is itself a fact. Where our politicians have said something demonstrably untrue we want a record, not to wipe the slate clean.

59. Information hygiene suggests that there is a pure information flow that should be flowing through our pipes like pure water, but someone has polluted it with salt. Or maybe they have polluted it with chlorine, for our own benefit.

60. Whether information has been polluted to the extent that it is no longer healthy is not straightforward.

61. If you take one of the laws I am concerned with, foreign interference, it is possible to commit this offence through circulating information if it is intended to have an interference effect, done on behalf of or to benefit a foreign power, and the information is misleading[9].

[8] 'CIA says lab leak most likely source of Covid outbreak' (BBC News, 26.1.25).
[9] Section 13 National Security Act 2023.

62. What does misleading mean? – under the statute it includes, where the misrepresentation is made as to a person's identity or purpose[10], so it would include astro-turfing campaigns. But while this may be a useful element of a criminal offence, just because someone tells a lie about their identity it doesn't mean the information itself is untrue.

63. There is also palpably false information that contains an element of truth and where, it could be said, that element of truth was particularly important to have within the information system so that overall it was healthy information, or closer to information than disinformation despite the element of falsehood.

64. Take the 'Trojan Horse letter' scandal. The findings of a government commissioned review in 2014 were both the letter was false but that there had been "a co-ordinated, deliberate and sustained action, carried out by a number of associated individuals, to introduce an intolerant and aggressive Islamic ethos into a few schools in Birmingham"[11]. Without the letter, this truth would not have been exposed. The letter contained an important element of truth.

65. I am not alone in not understanding terms like dis- or mis-information. OFCOM's 2024 report on public attitudes to misinformation (which it defined as false or misleading) found that the public considered misinformation to include not only the provision of empirically false information but also: the provision of information that someone doesn't agree with; and the provision of information that doesn't fit with someone's prior knowledge of, or existing beliefs about, a subject; and potentially something that a public figure said in fact said which was then accurately reported by a news platform or service, if the statement was in fact false[12].

[10] Section 15(6)(a) National Security Act 2023.
[11] Report into allegations concerning Birmingham schools arising from the 'Trojan Horse' letter (July 2014, HC 576).
[12] OFCOM, 'Understanding misinformation: an exploration of UK adults' behaviour and attitudes' (27.11.24).

66. Whatever you think of the term disinformation as a diagnosis, it is certainly not a basis for action.

67. Save in perhaps the most extreme circumstances (such as war), no democratic government has the authority or capacity to identify what information should be removed from circulation on the grounds that it is untrue. Attempts by politicians to redefine basic terms such as woman should lead to a great deal of humility when it comes to objective truth.

68. The free speech absolutists, whose views I do not share, are right with this observation: if you interfere with free speech then someone else is deciding what you can read. In other words, calling something disinformation does not provide policy solutions, and certainly not at scale.

69. It is impossible to craft an online safety duty based on a duty to remove disinformation that could capture the false conspiracy theories but not the true ones. Even more so when seeking to identify whether misleading material has been introduced or promoted from abroad, sometimes referred to as coordinated inauthentic behaviour.

70. In summary, disinformation and misinformation are not secure categories. And even if it is possible to distinguish healthy from polluted information, there are very strong arguments why these categories could not be the basis for government action save in the most extreme circumstances, such as during time of war.

71. As a final observation, I fear that dis and mis-information as being used as external factors to relieve us of our responsibility. The late Sir Alex Younger former Chief of MI6 observed that the Russians "did not create the things that divide us, we did that to ourselves."[13] It is a comforting myth that we are all really a nice set of people in the UK and we'd all get along so well if it wasn't for those pesky Russians.

[13] Obituary (Daily Telegraph, 3.6.26).

Free Speech Is Not the Only Game in Town

72. As I have already said, anyone who tries to use disinformation as a basis for action against the mind-messing horrors of the online domain, is likely to be met with robust and well-made free speech arguments.

73. The mistake which the free speech absolutists make is to think that this exhausts any arguments against government intervention.

74. It does not. It is true that my post on X or Instagram or in an online forum is, yes, a point of free expression. But what determines how far my expression reaches is different. This is clearest in the case of social media: it is the freedom of the tech company to promote my content to an unpredictable number of people based on a commercial algorithm. That has little to do, I suggest, with the free speech of the individual.

75. Even outside algorithmic promotion, there is another factor at work which cannot be ignored. It also demonstrates why free speech arguments do not adequately cover the field.

76. The journalist Janet Daley put her finger on this recently.[14] Her essential point was that today's extent of misleading information, doctored images, false news calling for social disorder has generally been associated with dictatorships not with free societies. She wrote:

> "There is a serious misunderstanding here of what constitutes free speech in a democracy. It does not stand alone, outside of any moral context or system of rules. The First Amendment to the US Constitution is usually cited as the purest expression of it, but it is a right embedded in a document which sets out a moral and political framework for the organisation of a society. It is a part of the social contract between

[14] 'We have misunderstood the meaning of free speech' (Daily Telegraph, 13.6.26).

government and people in which responsibilities are the price of guaranteed rights."

77. In her argument, what is different about social media is the lack of taking of responsibility. Tech platforms cannot be sued for libel. And those who post the vilest information, or circulate it, or purport to inform when they are instead trying to mislead, are too often anonymous.

78. She observes that, "You have a right to express an opinion, but you must identify yourself and be prepared to be held to account. It is the acceptance of responsibility for what is said that makes free speech possible. Without it, the liars and the authoritarians win the battle outright."

79. Of course the cover of anonymity is perfect for hostile states that want to stir up hatred against minorities by creating or promoting content.

80. But even without the intervention of hostile states, anonymity changes how people express themselves.

81. Martin Innes is one of many researchers who has drawn attention to the online normalisation of false persona and fabricated identity. He observes that when someone is able to obscure or obfuscate who they really are, "different forms of behaviour, often of the more problematic and troubling kinds, become more viable and plausible"[15].

82. There is a good reason, writes Janet Daley, "…why any publication which expresses political views, whether newspaper or campaign leaflet, must display the names of its publisher and printer."[16]

83. I agree with the UN Rapporteur on the promotion of freedom of opinion and expression that the effect of the free speech argument is that "political rhetoric

[15] 'The Deceptors' (Crest, 15.4.26). See also, Blumer, T., Döring, N., "Are we the same online? The expression of the five factor personality traits on the computer and the Internet", Cybersecurity 6(3) (December 2012).
[16] Supra.

stigmatizing and vilifying migrants, ethnic and religious minorities and other marginalized groups is on the rise around the world and increasingly being justified, even in some liberal democracies, as 'free speech'"[17].

84. It also follows that attempts to stop groups of people being stigmatised, such as Israeli and Jews, are denounced as 'censorship' or a violation of freedom of expression.

85. I also agree that the free speech argument fails to capture the fact that social media leads to convergence and conformity, and, in the case of pile-ons, timidity, so that aspects of online behaviour have the effect of dampening freedom of expression.

86. The classic free speech argument, that good speech drives out bad speech, has quite obviously been disproved by the evidence. Justice Brandeis, with whom that maxim is often associated, was writing for a pre-online age in which people were accountable for what they said[18].

87. There are counterarguments.

88. Firstly, there are some individuals who are willing to put their names to false statements, most notoriously President Trump whose lies are so brilliantly discussed by the legal and political philosopher Jeremy Waldron, also the author of a seminal book on counter-terrorism in the White House[19].

89. But in these cases there is at least the possibility of accountability. And who can tell whether politicians would feel as emboldened to peddle falsehoods if anonymity had not made it so acceptable?

90. Secondly, there could be whistleblowers or people whose ability to express themselves depends on not having their identities disclosed. I am sceptical about this. A whistleblower can use a secure messaging app. It must be

[17] 'Threats to freedom of expression online in turbulent times', A/80/341 (18.8.25).
[18] Whitney v California (1927).
[19] 'Damned Lies', Political Philosophy (2024) 1(1).

possible to create technical or social solutions for those who genuinely need anonymity when posting content in a public forum.

91. To sum up this part of my argument: whilst disinformation as a concept is vulnerable to a free speech argument, free speech should not be the only value system by which we measure regulatory changes to today's online environment.

Emotional manipulation as a better framework

92. In the concluding part of my analysis I will consider its wider implications.

93. To reiterate, I am not arguing that we should ignore or understate the impact of the online conduct that is discussed under the umbrella of disinformation. What I am saying is that these are not useful analytical tools.

94. To test this proposition, imagine that Parliament adds a new safety duty to the Online Safety Act 2023. They call it, "the disinformation duty of care". The duty is to minimise the likelihood that an online user will encounter disinformation on a service. Let us say disinformation is defined as factually inaccurate information about current or past events.

95. The first huge problem is that the service provider will not know whether something is disinformation at the time it is posted. Imagine the scenario: a user posts an image of a man attempting to behead another man on a Belfast street. What tools could the platform use to assess whether this was disinformation in the hours before an official police response?

96. Let us say that this image is being widely but organically circulated by users, and that this is contemporaneous with incidents of violence against migrants. Clamping down on accurate information merely because it is strongly associated with racist violence may appeal on the grounds of public safety but it is a censor's charter. It has nothing to do whether the information is true or not. If censorship is to happen – and there are cases where censorship is

justified – think for example terrorist livestreams or Islamic State propaganda – then it has to be justified on more precise grounds.

97. Secondly, if there is a Southport-type riot going on, and I circulate an accurate image of an Islamist Extremist attack from 5 years ago, insinuating but not quite saying that it has only just happened, our definition of disinformation will prove inadequate. The information is true but it has been presented out of context.

98. So let us adjust the definition of disinformation to include accurate information which is presented in a way that misleads. But whether something is implicitly misleading is very hard to judge. How is the platform to say – taking our Trojan Horse Letter example – that the essential feature of the information is misleading or not? What can platforms do about a conspiracy theory whose central message is so wild that it is impossible to disprove?

99. I suggest that the better analytical framing is emotional manipulation. The vice of digital media is that it makes us angry. In policy terms, our target should be the anger, not the information.

100. I think this leads us towards potential solutions.

101. I accept that anger is a valid emotion, and that anger can be a useful and galvanizing emotion. But when we signed up for the internet, did we really expect to be made angry just to keep us online so they could get more of our information and sell us things?

102. Did we anticipate that people would make careers off ragebait? Or that it would end up making us angry with one another?

103. Emotional manipulation is making our online lives less free. Flooded by emotion, the online domain is pulling us away from all the lovely things in the world, including from our fellow citizens, with whom we share a country and have so much in common, if only the rage would abate. We are the product and we are exhausted by this.

104. It interferes with all the wonderful benefits of the online domain.

105. Emotional manipulation makes us easy targets for foreign states who intervene in the online world to amplify our emotional grievances and that damages our national security.

106. So why not measure the effectiveness of an intervention by how much it deals with emotional manipulation? This is not about censorship or imposing a particular view of the facts. It does not stop people saying things online, but it might affect how far that content reaches.

107. So as a concrete starting proposal I would suggest removing anonymity for the mass emotional manipulators.

108. Take the X account, 'Inevitable West', with its 385,000 followers.

109. On 24 June this account, rumoured to be run by an Indian national, posted a picture of a teenager murdered in Narbonne in France by a "migrant gang". On 25 June, this account reposts a video of someone saying he had identified the attackers, with the message, "fear is gonna switch sides". That video has already got a million views. You can imagine what the comments are saying.

110. 'Inevitable West' is seeking to make me angry. It's not only that I feel I have the right to know who is trying to provoke me.

111. I also have two predictions. Firstly, if this account carried the name of a verified individual it would operate differently because of the possibility of being held to account. Secondly, knowing this individual's real name would allow media and civil society to investigate the account and its motivations. This would be a modest incursion into digital freedom but one that would be fully justified.

112. It would be recognising in the online context what has been known for a long time in the context of dirty money: anonymity is a problem.

113. A duty to limit largescale anonymous emotional manipulation under the Online Safety Act could either lead to a platform refusing anonymity all together or refusing anonymity for individuals who were engaged in large scale emotional manipulation. Alternatively, if anonymity was a central feature of its online offer, then platforms could alter its algorithm or the way it structured its financial incentives.

114. So to conclude, it is vital for academics and officials to consider the emotional aspects of online engagement. Disinformation is not a useful analytical or policy tool. I suggest there is an urgent need to develop moral and legal and definitional frameworks for when emotional manipulation is and is not acceptable.

JONATHAN HALL KC
2 July 2026