Generative AI, Drones, and Terrorism (London, 22.10.24)

Introduction

1. For what seems like far too long this year, I have been studying the risks and benefits of Generative AI for terrorism and counter-terrorism. Gen AI is the subset of artificial intelligence that produces original content in the form of text, audio, and imagery. I want to use GenAI to illustrate the risk of emerging tech in the terrorism domain.

2. We are most of us familiar with ChatGPT developed by OpenAI, Gemini by Google, Claude developed by Anthropic and so on; and if not with them with chatbots; and even if we are not deliberate users of GenAI, this technology is becoming increasingly integrated into word processing and predictive software.

3. As Prof Lewis Griffin of University College London has elegantly explained, the Large Language Models that lie behind GenAI operate on the simple principle – data in = intelligence out[1]. The more data the more intelligence, and the level of intelligence is awesome.

4. Policy makers refer to the GenAI lifecycle. This refers to all the stages involved in creating the computer models, sourcing data, training them, refining them, creating applications and interfaces, monetising them, hosting them, until eventually GenAI ends up in the hands of the consumer or end-user. The LLMs are at the top of the lifecycle – the end-user is right at the bottom. I will refer to the GenAI lifecycle later on in this speech.

Seven categories of GenAI terrorist harm

5. It is obvious that Generative AI can be used for wonderful insights, such as predicting the toxicity of drug compounds, and identifying novel cures for human diseases.

[1] "LLMs turn Data into Intelligence, like Steam Engines turn Fuel into Power" (2023) Large Language Models & Influence, DSTL Technical Report: TR149009 (access via the Athena depository).

6. It is also obvious that GenAI can threaten terrorist harm – a model that can be used to exclude toxicity can also be used, as researchers noted in 2022[2], to identify toxic compounds or develop vaccine-escaping Covid variants.

7. I have identified 7 categories of potential terrorist harm arising from GenAI,

    - Propaganda Productivity (e.g. easy translation into multiple languages of terrorist publications)
    - Propaganda Innovation (e.g. use of persuasive deepfakes)
    - Chatbot Radicalisation (one that concerns me most because laws against terrorist speech are designed with human speakers in mind)
    - Attack facilitation (practical on hostile reconnaissance or weapons design)
    - Attack innovation (toxin manufacturer or more plausibly cyber attack or use in an unmanned aerial system)
    - Moderation evasion (overwhelming content moderation systems used by tech companies)
    - Social degradation (use for disinformation, leading to more fractured and angry society where terrorism is more likely to take root).

8. I need to add 3 qualifications.

9. Firstly, I am acting on the as yet unproven assumption that particular terrorist content in the form of words and images causes terrorist harm in the real world. To my mind there are real world cases – for example, all those attacks explicitly inspired by the Christ Church footage - which support this assumption. But there are some who think this overstates the power of online content. And there are constitutions, most significantly in the US, which would always emphasize the merits of free speech over any of its negative consequences.

10. Secondly, we must acknowledge that GenAI could have unpredictably benign effects. For example, widespread permeation of deepfakes within the online environment could undermine propaganda campaigns based on authentic material. We know that images of atrocities distributed on social media played

[2] Phillips, D., et al, 'Generating Immune-aware SARS-CoV-2 Spike Proteins for Universal Vaccine Design', Proceedings of the 1st Workshop on Healthcare AI and COVID-19, ICML 2022; Urbina, F., Lentzos, F., Invernizzi, C., Ekins, S., 'Dual Use of Artificial Intelligence-powered Drug Discovery', Nat Mach Intell. 2022 Mar;4(3):189-191.

a major role in attracting Islamic State supporters to Syria[3]. In a world of deepfakes, videos of executions may be doubted and become less appealing.

11. Thirdly, as well as terrorist harm but there is the fact that GenAI is already being exploited by hostile state actors. In February 2024 OpenAI issued a threat alert warning that that 5 threat actors (affiliated variously to China, Iran, North Korea, and Russia) had used AI for researching intelligence agencies and satellites, debugging code for phishing attacks, and honing cyber attacks[4].

Framing the risk

12. What I want to do today is to consider the counter-terrorism response to GenAI by reference to society's responses to two other technologies that, like all technologies, can be deployed well or badly. I fear that some perspective can get lost in the rush to respond to GenAI. I will ask what insights we can generate from our response to these earlier technologies.

Two technologies

13. The **first** technology I want to discuss is the automobile. Cars are used by terrorists to kill people – for example down the road at Westminster Bridge in 2017. They can be used for terrorist reconnaissance and logistics. We have not banned cars, or called for a moratorium on their development and there is little moral panic about the availability of cars to terrorists.

14. It is something we live with. We don't go in for special warning signs (for example: "Caution: that parked car could belong to a terrorist") or resilience education in schools ("Beware if you are crossing a bridge with cars"). We accept the risk but go about our lives.

15. I suggest the key reason for our acceptance of this ubiquitous technology, despite our awareness that it can be used for terrorism, is market regulation. It is frankly not because of counter-terrorism laws that we accept the presence of cars.

[3] Stern, J., Berger, J.M., 'Thugs wanted – bring your own boots: how Isis attracts foreign fighters to its twisted utopia' (Guardian, 9.3.15)
[4] 'Disrupting malicious uses of AI by state-affiliated threat actors' (14.2.24).

16. Manufacturers want to sell cars. It is a mature market in which manufacturers and consumers understand need for regulation. All accept that there is a level of risk that means approvals at various stages, and safety obligations.

17. Consumer use of vehicles on the road is highly regulated. Driving tests. Driving licences. Systems linking vehicles to vehicle owners. A system of enforcement for breaches of the driving code. General acceptance that if you are going to produce something to go on a road or drive something on a road then it needs to be regulated.

18. So even though we know that automobiles can be exploited by terrorists, society feels a degree of comfort in the standards and expectations that exist. We don't panic about their occasional potential misuse. At most, there is selective target hardening - bridges in London for example - which stems from our experience of terrorist attacks and is a relatively minor adjustment to our way of life.

19. I suggest that this regulatory settlement is why the UK has already passed laws permitting the use of fully autonomous vehicles. The Automated Vehicles Act 2024 contains a regulatory framework, including criminal offences, to govern use of fully autonomous vehicles when they eventually come on the market and are approved.

20. The **second** technology I want to discuss is the internet. Every year in the UK more and more children are arrested for terrorism (42 last year) and I am confident in predicting that each one of these arrests was connected in one way or another to what the child has encountered online. I know this is not confined to the UK. Unlike automobiles, there is significant moral panic about the online experience because of its role in terrorism offending.

21. Unlike cars, the internet is not highly regulated. Being global, and being concerned with free expression, it is difficult to identify, let alone enforce national standards. Unless unacceptable controls were to be placed on free speech, you cannot regulate how people express themselves online in the same way as a driver can be required to stay in their lane.

22. But the early stages of the internet were characterised by a recklessness (for example, Meta's "Move fast and break things"). Harms were and still are ignored or minimised.

23. There is a thrill in inventing new ways of connecting us, creating new affordances, new ways of grabbing and keeping attention, irrespective of the cost to consumers. Companies do not want to have costs of moderation or

checking identities. They do not appreciate the scrutiny of civil society and the hard questions about how they operate. We don't know how some companies make money. It appears some of them just want to grow: hence Telegram's 900 million unique users and only 100 employees.

24. Unlike automobiles, there is moral panic. This is brought about not simply by the sense that there are few effective rules, but there is a lack of effective choice. As consumers we do not shop for apps like we shop for cars. We do not look at safety records. For example we don't routinely ask whether a particular app is suitable for our children by reference to its role in terrorism offending. We know the internet is simultaneously damaging our health and our children's health but don't know how to control our use or make smart choices.

25. Unlike the regulation of automobiles, society's response to the internet is fragmented and still at the experimental stage. Witness different responses in different jurisdictions, some of which (in Australia, Brazil, and France) have involved quite sharp confrontations between tech companies and national rules. The UK's Online Safety Act 2023 is shortly due to come into force which is yet another way of trying to limit the harm.

Like Cars or Like the Internet?

26. So I want to ask this question: Will our response to GenAI be more like our response to cars, or our response to the internet?

27. The answer I think depends on the GenAI lifecycle. You will recall at the top of the GenAI lifecycle are those individual and companies responsible for creating Large Language Models. At the bottom are the end-users, you and me. I accept that the higher up the GenAI lifecycle, the more that a car-type response is relevant; the further you go down the Gen AI lifecycle, the less like cars and the more like the internet it could end up being.

28. Let me explain why. Large Language Models at the top of the GenAI lifecycle are very expensive to create and require huge computing power, large amounts of data, and training and adjustment. There is also acceptance by the tech players responsible for creating this cutting-edge technology that GenAI brings major if not existential risks (for example, read Mustafa Suleyman's "The Coming Wave"; or listen to Sir Demis Hassabis, whose work with AI on the structure of protein has just won him a Nobel prize). Like car manufacturers they accept the need for product safety. So effort is put into so-

called “human alignment”, which includes avoiding the generation content that could be useful to terrorists.

29. Whether you do it through legislation – the EU’s AI Act – or voluntarily – the White House principles – the knowledge that there are some rules and values that apply to this powerful technology is a positive development; the alternative of banning, or freezing, this technology, because somewhere down the lifecycle someone may use it for terrorism purposes, is unthinkable.

30. As I’ve already said, as you descend further down the Gen AI lifecycle you are not dealing with big companies who accept the burden of standards but smaller players fighting for market share and ultimately billions of end users. So-called “guardrails” that are intended to prevent GenAI misuse can be modified by app creators or circumvented by end-users. Benign models can be trained on bad data – a particular risk as models become miniaturised. Generic off the shelf applications can be exploited. In January I created an Osama Bin Laden chatbot using a popular platform[5]. It was very easy to do and I suspect no amount of upstream regulation could stop me doing it.

31. If we are to avoid the moral panic that arises from terrorist and other harmful use of the internet, I suggest that we need to avoid the errors of the early internet period, and start setting standards now. It is not just resilience for when things go wrong, but a sense that GenAI is held within society’s overall rules and standards.

32. Lawmakers need to demonstrate to the public that GenAI is not a lawless zone. It must be understood by the law, it must be understood by law enforcement, it must be understood by civil society. We must not make the mistakes of the early internet period of believing in tech-utopianism, or that any problems left to the tech companies.

33. That means, for example, ensuring terrorism laws and laws against racial and religious hatred are kept up to date. It means reacting to clear use-cases where GenAI is being exploited, and being prepared to develop new laws at speed, for example against harmful precursor behaviour. Even if we don’t know precisely how GenAI is going to be exploited by terrorists, we need a common understanding of GenAI, and a confidence to act, and certainly not a reaction that says, this is just “too difficult”.

Drones

[5] Hall, J., ‘Al-Adna did not stint in his glorification of Islamic State’ (Telegraph, 1.1.24).

34. I'll turn briefly to Unmanned Aircraft Systems or drones. Again, Drones offer wonder opportunities but can be abused. We are aware of drones being developed for battlefield use by terrorists. Drones have been adopted by criminals – for example, to fly contraband into prisons – and will no doubt be adopted further by terrorists.

35. In UK there is regulation affecting the use of drones and it includes some provision for licencing users, but the regulation does not apply to 'toy' drones. If we were just dealing with high-end expensive drones – the drones that can carry payloads in the 10s of kilograms – then the car model might work. In principle one could consider regulations on manufacturers, geo-blocking on sensitive sites, user licences.

36. But as prices come down and consumer appetite increases, we may be dealing more at the toy end of the market. Manufacturers may race out new models without thinking of consequences. If the law does not keep pace with technologies, then moral panic will set in. What will we do if drones are raced down streets or used to harass traffic? Will we accept an anarchic freedom as we have with the online world? I suggest that in the absence of laws, then terrorists will be emboldened to use drones; and society may end up overreacting, and miss out on their benefits.

37. We need to work out whether drones are more like cars or more like the internet. I suggest that regulating not just manufacture but use of drones – high-speed moving objects – is much more akin to the car model where anonymity is not tolerated and rules are widely applied and accepted. I suggest that going down the internet model is not an option.

Conclusion

38. In conclusion, let us consider the car model; and the internet model.

39. As far as possible, we need a settled environment where GenAI is welcomed as part of society and governed by rules, even though there will inevitably be misuse.

40. On the other hand, we need to avoid the mistakes of the early internet period where there is a free for all leading to a sense of moral panic. I think we can achieve this, so long as we put an understanding of GenAI at heart of legal and policy decision-making. Thank you.

JONATHAN HALL KC
INDEPENDENT REVIEWER OF TERRORISM LEGISLATION