Edward Snowden on Palestine Artificial Intelligence and Individual Agency

Edward Snowden

ChatGPT *Substantively cleaned English transcript*

Editorial note. *Timestamps, stage directions, the technical interruption, filler words, false starts and repeated formulations have been removed. Recognition errors and sentence fragments have been corrected while preserving the speaker's argument, examples and rhetorical character.*

Opening remarks

Edward Snowden

Everybody can hear me, right? I have to check. All right. So I have to say I haven't been doing public speaking in a number of years. I might be a little bit rusty. I have become a parent since. Instead of authoring tweets that felt as if nobody was paying attention, I took a step back. The problem, as I think you probably understand in your own lives, is that everybody is yelling and nobody is listening. There's not much conversation happening anymore. AI certainly hasn't helped that. Stepping away and entering family life has helped me look at things again through the way children understand them. I then have to recognize the slightly tyrannical parent within myself who says, 'No, it is dinner time.' No, you have to stop what you're doing right now and actually impose the rules. And that's an unnatural state of being for me.

But it's been quite instructional because ultimately somebody has to do that. But what this means is that I've been spending a lot of time reading fables. Two nights ago, while thinking about what I wanted to say tonight, I came across a story you may recognize.

The Emperor's New Clothes

The emperor's new clothes. Remember that one? It has probably been a while, so I will go through it. Once upon a time, con men arrive in a kingdom ruled by a famously vain emperor who is preoccupied with appearances. He squanders the treasury on fancy clothes and insists that no one be more beautiful or interesting than he is. The con men see an opportunity and claim to possess a magic fabric visible only to the best people. Anyone who is unwise or unsuited to office, they say, cannot see it. Despite his vanity, the emperor still has some interest in governing. He thinks the fabric will reveal who is unfit for office and sends his most trusted minister to inspect it and commission a suit. Everyone hears about the wonderful, extraordinarily expensive clothes. The con men make a great show of working at a loom, weaving fabric that no one can see. So, what happens?

When the minister inspects the work, the con men ask what he thinks. Everyone knows that anyone who admits being unable to see the fabric will be judged an idiot and unfit for office. The minister therefore adapts instantly to the lie and declares that the emperor will be delighted. The finance minister, foreign minister and other officials follow. One at a time, without pre-coordination, conspiracy or agreement, they see the others pretending to see the fabric and instantly know what to do. Each claims to see it and praises it. You know how the story proceeds. The story ends when the emperor, admiring his new clothes, steps out into the street. All the townspeople - the respectable, civilized adults - see him walking around naked but praise his dazzling raiment. Then a child, ignorant of the socially acceptable lie the adults tell themselves, points out that

Gaza and the socially acceptable lie

There is in fact a genocide occurring in Gaza that its exterminationist intent has been openly avowed by a number of Israeli ministers both on camera and in writing and that for western democracies to continue arming such an operation given the totality of evidence is not only unethical, but it's most probably criminal. And there's actually, in fact, nothing antisemitic or even anti-Israeli in saying that. We should expect better. We established international law together during the same historical period. We all had a hand in making these rules and so we should abide by them. Children are often the ones who recognize that we can live according to our ideals rather than simply accepting reality as it is.

Personal oaths and integrity

The moral is that there may come a moment when everyone else in your office, laboratory, city, state or country is wrong. They may know they're wrong, they may not. We build up these big and complex systems of law and belief and behavior and we all want to be good and decent people who do the right thing. But sometimes the requirement to fit in conflicts with the requirement to behave humanely. What do you do when the moral is in conflict with the legal? In my case, I faced a conflict between my oaths. That is why some of you know or recognize me. I struggled with this inside. I talked to other people in my workplace. I watched the U.S. Director of National Intelligence, General James Clapper, swear to tell the truth before Congress. A senator on the oversight committee asked him directly whether the NSA was collecting data on millions or hundreds of millions of Americans.

Clapper immediately realized the difficulty of his position. He rubbed his head, began to sweat and answered: 'No, sir. Not wittingly. There may be cases inadvertently, perhaps, but not wittingly.' Meanwhile, I was effectively the IT guy, a systems administrator in the Office of Information Sharing. That was what they hired me to do. They did not realize how good I would be at that job. I had sworn an oath to defend the Constitution against all enemies, foreign and domestic, but I had also signed Standard Form 312, the classified-information nondisclosure agreement. It says you're not going to tell anybody about what it is you do. So people can disagree about the hierarchy between these two documents of course but in my case I felt that the constitution was owed a little bit more weight. That actually turned out to be a somewhat controversial decision.

But there are moments in life like that right and you have to decide what to do and how you're going to go about doing it and how much to risk. One thing I strongly dislike is being introduced in a way that turns me into a heroic figure. And I want you to understand that I'm not a hero and I don't believe in heroes. I'm not a perfect person. I do terrible things. And sometimes I feel bad about them and sometimes I don't feel bad about them. I am human chiefly. Frankly, it's been a struggle to get to that point. But that's where I'm at now.

Whistleblowing and truth

When we call someone a hero, we distance that person from ourselves. One of my personal heroes is whistleblower Daniel Ellsberg, who revealed the Pentagon Papers in the 1970s and was called 'the most dangerous man in America' by the president's circle. You think they did something exceptional. You think they did something that you can't do, but you can, right? If the IT guy, if the guy who fixes your printer can do it, you can too. And frankly, guys, I didn't graduate from university. Anybody sitting in this room in Zurich is probably a lot further down the road of intellectual development than I was at the time, even at 29. I was a little bit younger. Obviously, I've got a few more years on me now, and I think I understand things a little bit better than I did back then, but generally, I think I made the right call. I might do it a little bit differently.

I frankly might have done it a little bit more aggressively than I did back then, but I wanted to be overcautious. I wanted to be more careful. I wanted to trust journalists to check my instincts, determine the public interest, and balance my beliefs against their judgment and the government's arguments. That is becoming increasingly difficult because there is a new predilection for even western democracies to engage in information control that frankly came

out during the COVID era when questioning lockdown policies and mandates was treated as dangerous to the public. You can understand the concern, but the approach taken was fundamentally at odds with the ideals of an open society. And I think in a lot of ways it's gotten worse. We spend a lot of time now talking about misinformation. We talk about disinformation.

We have Western officials who are very blatantly lying to the public, which is not a new thing. That's happened in every government throughout time. But now we see journalists sort of going along with it. Not because they believe them. The journalists definitely don't believe them. They have the savvy. They have the understanding. They have the background to understand. They go along because they regard doing so as prosocial, and that is dangerous. And I want to get to the real point of where I'm going at tonight because we've got a lot of things to cover and we'll get to the panel. We can discuss a little bit more.

AI and the centralization of power

Artificial intelligence, particularly the large language models through which ordinary people experience it today, has changed our understanding because it centralizes power in the same way, but on a much greater scale. We used to endlessly panic about the authority of Google and how they controlled things like search results. But a lot of people aren't even using search anymore. They're just asking the LLM. The behavioral standards embedded in these systems - how they interpret law and social organization - create contradictions that your generation will have to resolve. I am out of the workforce now; the CIA and NSA no longer see fit to employ me. The introduction described me as a contractor, which is correct for my final position in the Office of Information Sharing.

Originally, however, I was a CIA field officer and direct government employee, working undercover in Geneva and elsewhere. The distinction between government and private-company employees began to blur in the United States in the post-9/11 era. And that distinction has only become more difficult to differentiate as corporations have accrued more and more power and we start to see these Palantir-type arrangements in which government has access to a data repository but lacks, or does not want, the personnel required to build, operate and oversee everything in-house. That's a lot of bureaucracy. They would rather spend their budget in, different ways, however they want to do it. Instead, they place the personnel burden and much of the operational responsibility on private companies, asking them to construct and maintain the government's data infrastructure.

Private companies as agents of government

Private companies are operating the infrastructure and becoming government partners - in a sense, agents of government - without actually being government employees. 'Agent of government' is an important term in U.S. Law. Ordinarily, the Constitution does not bind private companies. Google doesn't have to respect your freedom of speech or association or anything else like that. The Constitution binds the state, except that under the state-action doctrine it may also bind a private party directed to act as a government agent. It's very rare to see this applied. The important point is that someone must bear responsibility for the action. The government can't just point to the company. The company can't just point to the government and have it go away.

AI liability and human responsibility

Well, that's what we're doing with LLMs. There is much discussion of OpenAI and Anthropic, which have developed advanced models and continually leapfrog their competitors. But they did this by scraping the commons. They trained them on my book, as emerged in a lawsuit, as well as my other writings. And everyone else is out there, right? Like I'm not special. This is anybody who's commented on an internet forum ever. Anyone who's been engaged in these things. They trawled the corpus, they distilled the commons, and they trained a model off of that. And actually, I'm not bothered by that. I think that is broadly acceptable. They violated the copyright, and I want them punished for that, mostly just because I want to hinder them enough that everyone else can catch up more so than I care about the fact that they're, studying from my notes or whatever.

But the idea is that now they're talking about agentic AI which I think is a huge mistake that they're going to be discovering very soon. They should not brand this technology as your agent because if the government uses it, you are an agent of government and that means the constitution applies to you. If a government uses a model running on a company's server, the company receives the queries, applies its system prompt, recomposes them and runs them through its systems. If the model makes an error - or even functions as intended - people may die as a result, as reporting about strikes on the wrong buildings in Iran suggests. Because oopsie, the LLM made a mistake. And tremendous numbers of, children, the most innocent that we can imagine are being killed. Well, who owns the liability there?

And of course, naturally, they would say the government owns the liability because you're never going to successfully sue the government for this, even though they should eat it. But maybe that doesn't work anymore because maybe they are an agent of government. And I don't know where we're ever going to see standing to sue on this, but we should. And people in this room, I think, are frankly going to be called upon in the course of their careers, in the course of their studies, in the course of whatever you do to contend with these kind of issues. Where are we going to draw a line? An algorithm cannot make a mistake in the same moral sense that a person can. Its erroneous outputs are not the computer's responsibility; the computer is a slave to its instructions. Responsibility lies with its developer, programmer, creator or architect. In other words, responsibility lies with a person.

And we cannot hide that. We cannot abstract that away. Right? A gun maker is not liable, it's true, for the misuse of the gun. But if during the testing of their latest piece of artillery they mistakenly obliterate, a nearby village, it's not the cannon that should go to jail. And just because you put out a press release describing how surprised you were that your new model, hacked the neighbor, or how impressed you are by the destructive power of your missile. That does not release you from liability for launching it. The missile did not launch itself; you launched it. And so there's a lot we can get into, but the idea that I want people to contend with is how we deal with this. And right now, these closed models are saying AI safety, AI safety, safety as they call it, where the idea is these things are so dangerous. Oh gosh, they're so dangerous, but they're so dazzling.

They're so important. They're so valuable. We can't get rid of them. But, oh, it's going to kill some people, but it's all right. That just happens. Oh, it's going to hack some companies. Oops. Whatever. Will invest in them.

Local ownership of devices

The central question is who operates the system, and the same principle should apply more broadly to our devices. Right now, you buy a cell phone from Apple and before you're allowed to use it, you have to click okay to continue and they go, well, you're assuming liability for that. You clicked okay to continue. Google sells you an Android phone, Samsung sells you an Android phone. You click okay to continue and now that's on you. That cannot reasonably be how responsibility works because you can't inspect the system. You can't understand how it's used and particularly in the context of LLMs where you're submitting queries to a model that they are running The provider intercepts and recomposes the prompt, applying undisclosed safety rules and transformations. Responsibility for that cannot rest entirely with me when I cannot even see what the provider did.

I asked you a question, you gave me something back in return. And if it's an agentic AI where I'm putting this thing in a harness, right? The harness being just a little program that basically uses these large models to feed it instructions that it can then provide to tools that are existing on your local system. So now your computer that is connected to the internet is being frankly remotely controlled by a model that's far away that you don't understand how it works, but you hope it works pretty well. Frankly, they do work reasonably well, but every now and then it might hack, some random company or whatever out there. I think we need to put Sam Altman in jail, and I think it'll be an instructive experience. I think it'll be an instructive experience for the industry and for the world. And it wins me friends in this audience I appreciate, but it might not throughout the world.

I have not been following the public mood recently. But the idea is that it doesn't have to be like this. There is an alternative model: publish the weights and run large language models as open-source software and we see this

through competitors. They're still imperfect. With locally run, open models, companies no longer have to play hall monitor between the user and the output. When computation is pushed to the edge and the user truly operates the device - owning it, controlling it and inspecting what it does - there is a much fairer legal argument that the user is responsible. And that's the way it should be. If you are buying something, it is yours. It should work for you. And you guys are going to be involved in the architecting decisions about how future paradigms of people's interaction with software actually works. And I would like you to remember this.

A first law of technology

We need a kind of first law of technology, something like this. A device should never act contrary to its operator's interests. Safety matters, but if you instruct a computer to do X, it should not do Y while telling a third party that it did X. A device should serve the operator's will - ideally also the owner's will, assuming they are the same person. Leasing and divided ownership complicate the question, but they also reveal the underlying problem. That illustrates the problem. If a device is sitting in your pocket that tracks you, that understands you, that interprets you, that connects you to the world, that interjects itself into all of these different things, that intermediates everything.

It mediates everything and you depend on it to live, to work, to eat, to engage, to travel, to buy tickets, to establish yourself, to prove yourself, to permission yourself in the world, this must be yours. And if it's not yours, if it's somebody else's and you can't engage them, you can't travel, you can't establish who you are, you can't spend, you can't trade, you can't interact, you can't speak, you can't publish without these things. Well, that says something very dangerous about where sort of the allocation of power in society has become. I know I have gone over time, so I will bring this to a close. But the idea is that if OpenAI or Anthropic were trying to say they're for public safety and they're interposing themselves between your will and the actual output and they are transforming these things. They are saying that they know better than you how to operate your technology.

And perhaps that is true. But that is a very dangerous proposition that they are knowingly entering into millions of times a day, including in these like military, paramilitary, governmental contexts. Where if you were a passenger in someone's car and you go, "Oh, the driver's acting somewhat unsafe." and you grab the steering wheel and you pull it and you in fact cause an accident rather than avoid an accident. If the police show up at the scene of the accident, we go, "Well, well, I only pulled the wheel a little bit." The liability is still fully yours, and I think it should be theirs as well.

Law judgment and the duty to act

The point I want to leave with this audience is that many of these questions do not have clear, bright-line answers. As in 'The Emperor's New Clothes,' there may be a clear and socially expected answer, but that does not make it correct. And a lot of times we in developed societies like to substitute the law for our own judgment. We go it's the law. It's the law. Slavery was legal, right? It was wrong. We need to understand that the law doesn't protect us. We protect the law, right? We invest the law with meaning through our acts through upholding sort of these legal traditions. When the law begins acting in ways that are contrary to the protection of people and society, the law needs to change. If people can't see that, it is sort of like the morning bell. Someone must strike it to wake us, bring us to attention, right?

The spirit of the law might itself be slumber, but when it hears the bell, everyone pays attention. Now, that doesn't mean everything gets better. That doesn't mean you save the world. God knows. I tried to tell people something. And frankly, a lot more people heard it than I ever expected, but that didn't fix the problem. It's still out there. It's continuing. It's getting worse in many different ways. But it also got better in other ways. And when people have awareness, when we're aware of the danger, we can begin to steer around it. We can begin to correct. And that's a long process. It might take longer than our lifetimes. But I hate to say it, but that's kind of normal. Information by itself is not power, no matter how often we repeat the slogan. Organization is power. Our will and willingness to act are power. The information does nothing by itself. It can even pacify.

It can disempower people. It can sedate people. We see the doomerism. We see the endless scrolling. You see the atrocity in the newsfeed. You click the play button and then go, "Oh, I didn't want to see that." You scroll along without saying anything. You go to the next thing that makes you feel good. Self-certain in the illusion that you couldn't do anything about it. It feels inevitable. We have no time to act in a lot of cases. And people want to feel like they can't do anything, but you can. You have to look for the place in your own environment or office where everyone else may be wrong. And most of the times it's you who's wrong. Maybe I was wrong. And again that is normal. We try and we fail. We start out poorly. We are constantly making mistakes as people but we can get better. I think humanity in a large way is about failing.

It's about failing well, failing by less, failing by smaller degrees, failing better, failing big and striving farther. Failing deeper and faster until you've gathered such skill and confidence in correcting your errors and in facing the fall, the uncertainty that you can begin to bend the path of your decline into a launch, and suddenly now you're moving. Headlong dives accelerate and they build upon themselves with an arc and dignity like a bird, for whom the tyranny of gravity no longer holds fear.

Conclusion

That is the idea I want you to remember. We are not born human, but we do have the chance to become it through the choices that we make every day. Thank you.

Version 8.11.2026

Address of this page

Home

Joachim Gruber