

The Coming Tsunami of Falsehood

Daniel Hardman — June 2023

I'm writing this letter to forewarn you about a tidal wave that I see on the horizon. It consists of problems with truth that are big and scary, and I fear it could well drown our society. I am sensitized to this particular event because of my professional background in tech, though that's not the only viewpoint that could shed light on the topic.¹

About a decade ago, I shifted the focus of my career to cybersecurity. This is an interesting field, but it can be depressing, because the quantity and depth of digital dishonesty is mind-boggling. The whole industry would vanish if everyone told the truth.

I spoke at the RSA conference in San Francisco in 2016, and left with a feeling that 40,000 smart people (the tip of an industry iceberg worth tens of billions of dollars a year) were wasting their time. Sure, we made useful point solutions — but we were little boys with fingers in a leaky dike. And a big tide was coming in.

Maybe my pessimism at that point was not rooted purely in my profession. Political dysfunction, fake news, and reckless, manipulative (social) media were also having a huge effect on us, already.

Soon after that conference, I found hope in a totally new approach to how we manage our digital identities. Imagine the seismic shift in trust if people and companies logged in to you instead of you logging in to them...

I've pursued that idea ever since. The tech behind it is important and robust and ethical, and there are some bright spots in its adoption. However, I have also observed with dismay how existing political and economic forces militate against deep change. And the climate outside my professional activities continues to worsen.

Recently, big advances in artificial intelligence (AI) have made headlines. What just happened is that AI got WAY more data (like, millions of times more), it got creative (it can now generate hyper-realistic content of all types), and its ability to learn and reason

¹ I encourage you to take a minute to ponder *me* as a source of truth, because evaluating sources is an important protection. You should filter what I say through some healthy skepticism: I see in part, and I analyze from a particular lens. On the other hand, giving reasonable caveats is part of being trustworthy — and I am not a lone voice. Much of what I will say here is being predicted by numerous thinkers from varied backgrounds. I will try to cite some in important places. Bad sources tend to assert that they've already done all the thinking for you, and they often cite nobody (or cite only other bad sources)...

got much, much stronger. A lot of experts are predicting disaster from this new AI tech run amok. [Here](#) is one video that I already shared that summarizes a lot of it in a way that I find credible.

As a former full-time researcher in machine learning,² I feel moderately confident about distinguishing hyperbole from real problems in the field. Some pundits seem over-the-top to me — I don't believe a super-intelligent AI will evolve from our current systems, in months, and decide to exterminate humans like obnoxious bacteria. However, other predictions don't seem far-fetched at all, and they have me very, very worried. I think AI that ISN'T superintelligent and ISN'T conscious/self-aware but that IS wielded as a weapon by conspiring people is *still* about to make society's problems with truth much, much more dangerous.

Perhaps I should say it another way: *Next-gen AI isn't the real problem: dishonest people are. But if you combine the two, things get scary.* Liars have just discovered the information equivalent of nuclear weapons, and I worry that we won't have the discipline and wisdom to double down on mechanisms capable of reining them in.

Let's go back to my metaphor about little boys with fingers in the dike. In 2016, I thought the tide was rising, and I was dismayed. Since then I've observed a perfect storm on top of the tide: political demagoguery from all ideologies; discredited and dysfunctional government institutions; personal scandals of all kinds; social and traditional media operating with poisonous business models and insufficient guard rails; the entrenched, manipulative ethos of surveillance capitalism; economic headwinds and financial instability; terrorism; pervasive distrust of one another, along with growing isolation into curated digital (un)realities; tyrants and failed democracies; serious health concerns weakening the social contract everywhere.

So, that's the tide and the storm. Now add to that a *tsunami* of falsehood, and the dikes seem doomed. To me, this sounds like exactly the conditions that Russell M. Nelson predicted when he [said in 2018](#): "If we are to have any hope of sifting through the myriad of voices and the philosophies of men that attack truth, we must learn to receive revelation... in coming days, it will not be possible to survive spiritually without the guiding, directing, comforting, and constant influence of the Holy Ghost." That's a spiritual statement by a man I consider a *bona fide* prophet, five years before the AI breakthrough — but even if you just map it to a well-educated man who's seen a lot in a century of living, I think it's sobering.

² Machine learning is a subdiscipline within AI research. I am not a deep AI expert, and I haven't invented any AIs. However, I did maintain a sophisticated AI, fed by several billion data points a day, for two years. I learned a fair amount about the underlying concepts.

Here are some things that I think we may see very soon, enabled by the latest AI innovations:

- Scams and cyberattacks will combine deep knowledge about you (which they'll have from surveillance capitalism, social media overdisclosure, and the dark web) with AI. Instead of a scam or an attack that plays the odds to catch a few gullible people with a big net, the AI will, on autopilot, write a script that plays expertly on the aspects of your specific life where you are most vulnerable and easy to convince. The AI will generate a perfectly faked voice and video to accompany the script. For example, a parent will receive a phone call from an incoherent and terrified child, announcing that they've been kidnapped. Or you'll get a text from your teenage grandchild asking for emergency money. Or... The outreach will be full of details that we imagine nobody could know, like the location of your child's field trip today at school. [[example](#)]
- Using the same personal targeting technology, and combining it with fake identities, AIs will be used to craft individualized persuasive emails, texts, or posts on social media — pretending to be from a trusted friend — that are designed to manipulate the target. This could be used for political persuasion or selling, but also for far darker things: undermining a marriage, destroying someone's religious faith, feeding an addiction, triggering a suicide.
- Armies of AIs will be unleashed to do cyberbullying and persecution of the vulnerable on a scale and with a degree of laser-targeted malice never seen before.
- AIs will be used to create designer versions of street drugs (or to modify prescription drugs strategically) to make them more addictive, cheaper to produce, or harder to detect as contraband.
- The character of good people will be expertly assassinated with deep fakes of all kinds. [[possible example?](#) | [definite examples](#)]
- Major world events — both contemporary and historical — will be faked. We've already seen this with elections, but I think this is just the beginning. The ability to fake an unfolding world event ("Oh no! The Golden Gate Bridge just collapsed!") to create mass hysteria or to overload fragile resources and supply chains will be used as a weapon of aggressive states. [[examples](#)]
- People will begin using AIs to find weaknesses in software and loopholes in contracts, study how alibis can be undermined, groom victims, identify children who are easy to kidnap, discover which teens are vulnerable to sextortion, target

people most likely to be addicted if offered free drugs, optimize the effects of a terror attack on a supply chains, etc. [[example](#) | [example](#)]

- Any dating that's not in person will be suspect, because Tik-Tok filters and AI will allow someone to project an artificial picture of their face, voice, and body — even in real time Zoom meetings. Paradoxically, in-person interactions will become scarcer. Pornography will grow darker because AI will be able to paste anybody's face and voice onto any porn template, and will also be able to cheaply and instantly customize the scripts to suit anyone's fantasy on the spur of the moment. This will combine with realtime gaming engines, fully lifelike sex robots, and AI-powered adult chatrooms to allow a lonely person to live in a porn-saturated unreality if they want. [[discussion of this possibility](#) | [early reported examples](#)]
- AI will be used to perform priestly functions like hearing a confession, counseling the sinner, and giving sermons and church lessons. [[speculation](#) | [example](#)]
- AIs will be used to lure and bait teachers, business leaders, public servants, religious leaders, and missionaries.
- Organized attempts to discredit or undermine religion or scholarly work using AI will be a thing. Imagine Mark Hoffman, but with the combined knowledge of everything ever written on a subject in all known languages; a near-infinite ability to generate digital evidence saying whatever is strategically useful, in the voice or handwriting of anyone, on demand; vast citations; fake scholarly opinions which then intimidate other voices and manufacture a groupthink scholarly consensus; fake historical events; hard-core and targeted data manipulation to poison linguistic analysis and other statistical measures; character assassination of the living and the dead.
- Fake religions and fake charities will abound for the same reason, and many fringe elements will deepen their descent into quasi-religious movements because AIs will be used to analyze and manipulate the credulous and careless in ways that optimize revenue and reinforce foolish beliefs. This will further discredit real religions and real charities.
- Education and life/career experience will be trivialized as people regurgitate plausible things that AIs generated, as if coming from a human's hard work or insight. [I already saw an example of this in my career — a coder submitted the product of an AI as if it were their own original research.]

I will stop my list of depressing predictions there. Lots more could be added, but you get the point.

Phenomena like these could be compared to an upheaval on the ocean floor that starts a tidal wave. The tsunami itself is a massive surge in victims, damaged trust, and further erosion of the social contract. Society may be ripped to shreds. We won't know who to trust, we will feel abused, and we will be suspicious of everyone, with good reason. I think the divisiveness that we saw after George Floyd is just a tiny foretaste of the effects this tsunami will have on society.

If prospects are so bad, why I am writing this essay?

Because, although I think we will all be impacted, I also think we are not powerless. And I want to ask you to engage with me in a conversation about constructive actions. Here are a few that I believe in:

1. **We need to proactively seize the opportunity and responsibility of being careful truth-hearers and careful truth-tellers.**

Fifty or even twenty years ago, consuming information from reputable-looking sources would (often) lead to a reasonably accurate view of reality. Knowledge about cybersecurity and scams and fraud wasn't a crucial survival skill. Common sense might have got you by.

In the world I foresee, our standard of education and behavior on these issues has to be higher. We don't all have to be cybersecurity and AI and fraud experts, but we all need to invest in reasonable awareness and learn [some good habits](#).

We won't be on safe ground if we assume what we hear is fact-checked; we need to learn to do fact-checking ourselves ([snopes.com](#) and similar services are our friend). We won't be on safe ground if we assume that something is true because one source we like said so; we may need to triangulate. (ESPECIALLY, fact-check AIs! See [this doc](#) for a sobering example of why.)

Developing [a humble and careful and tentative attitude about truth](#) will stand us in good stead. [Being thoughtful about how we use AI](#) and other digital tools ourselves is ethically imperative, and it will also sensitize us to how those tools might be used against us.

2. **We need to get in the habit of proving the authenticity of what we say and do — and expecting others to do the same.**

Today, most of our communication has only an iffy connection to a source. How does a friend know which account with our name on it is *really* us on Instagram? Did we *really* say that on Facebook, or was our account hacked? If it was hacked, do we take seriously our responsibility to fix it? For that matter, how can a friend know it's *really* us (not a hacker trying to undermine our reputation) when they get a message saying our account was hacked?

Our reputations are our asset and our responsibility. Others may try to abuse them, but we can and must do things to safeguard them; nobody else will do it for us. How can we expect people to trust us, if we don't treat their trust as sacred?

Perhaps we need to think about having witnesses in contexts where we have been casual until now. We recognize this as an important safety measure in school and ecclesiastical settings, but maybe we need to apply it in our personal and work lives in new ways...

If we know with 100% confidence the origin of any communication (email, text, phone call, letter, tweet...), then we can make better decisions about whether to trust it. We should generate evidence about our own communication to give others that same benefit. Doing this is fundamental and profound, and I work with technology that makes this possible. I'll share more about that later.

3. **We need to get in the habit, *now*, of treasuring the best sources.**

This means we need to recognize which of our relationships and communities are characterized by careful and humble truth-telling, and which are not. We need to nurture the relationships that give us truth, and prize them — and de-emphasize ones that are iffy. Good sources will anchor us. As President Nelson said, we need a direct line to heavenly guidance that we know well and use continuously. He's wise and inspired, and he's not the only leader from the only religion to make such a point. It's hard to overemphasize this.

4. **We need to be clear to ourselves and others about what we (already) know to be true.**

This doesn't mean we should shout our prejudices to each other even louder than we already do. It means that we should develop the habit of looking into our own hearts and deciding whether and how we consider something to be

true, and holding ourselves to a high standard as we do. Then we need to speak more about those things, and less about things that our hearts trust less.

If we do trust something or someone important, we should be able to articulate why, and I suggest we should write it down. It's going to be challenged, and we will need anchors to ground us to such things as the tsunami presses upon us. Others may need our words, too. Capturing them now is better than later.

Please help me spread the word and figure out how to make things better.