



## Artificial Intelligence, Defamation, and New Speech Frontiers

Thursday, June 8, 2023

Visit our media library at [constitutioncenter.org/medialibrary](https://constitutioncenter.org/medialibrary) to see a list of resources mentioned throughout this program, listen to previous episodes, and more.

**[00:00:00] Jeffrey Rosen:** Earlier this week the first ever AI defamation lawsuit was filed by a Georgia radio host who claims that ChatGPT falsely accused him of embezzling money. It's time for a conversation about AI and the First Amendment. What happens if AI makes up false and damaging information about a real person? Should they be able to sue for defamation and who or what is liable? And how does AI which has neither reason nor conscience challenge the underlying principles of the First Amendment that we protect free speech to protect freedom of conscience and the rule of reason.

**[00:00:39] Jeffrey Rosen:** Hello, friends. I'm Jeffrey Rosen, President and CEO of the National Constitution Center and welcome to *We the People*, a weekly show of constitutional debate. The National Constitution Center is a non-partisan non-profit chartered by Congress to increase awareness and understanding of the Constitution among the American people. Joining us today to help us understand the emerging legal issues surrounding artificial intelligence and the First Amendment are the two leading thinkers on this topic. I'm so excited to convene them and can't wait to learn from them.

**[00:01:09] Jeffrey Rosen:** Eugene Volokh is the Gary T. Schwartz distinguished professor of law at the UCLA School of Law. He's the founder of the legal blog *The Volokh Conspiracy* and author of a new paper on AI legal issues called "Large Libel Models." Eugene Volokh, welcome back to *We the People*.

**[00:01:26] Eugene Volokh:** Thanks for having me.

**[00:01:28] Jeffrey Rosen:** And Lyriisa Lidsky is the Raymond and Miriam Ehrlich Chair in US Constitutional Law at the University of Florida Law School. She's a leading voice on free speech online and the author of several important articles on the law of defamation. Professor Lidsky, it's wonderful to welcome you to *We the People*.

**[00:01:42] Lyriisa Lidsky:** Thank you.

**[00:01:44] Jeffrey Rosen:** Eugene Volokh, let's begin by describing this very well time for our purpose Georgia case. Tell us what the, what, what happened and what the legal plans are?

**[00:01:54] Eugene Volokh:** Sure. So Fred Riehl, who is an online political commentator journalist. I think he runs the ammoland.com site which deals with Second Amendment issues was interested in a particular a Second Amendment related case that was filed in federal court in Washington. So he goes to ChatGPT and asks it to summarize the complaint in that case.

**[00:02:19] Eugene Volokh:** I'm told that actually ChatGPT is often pretty good at that, so it's unsurprising that, that he might try it out, because it's a long complaint. It's good to get a quick summary. But ChatGPT responds by saying it's a legal complaint filed against one Mark Walters who was accused of defrauding and embezzling funds from the Second Amendment Foundation. Second Amendment Foundation is indeed the plaintiff in that case, but Mark Walters has nothing whatever to do with the case. He, he is himself a commentator on Second Amendment issues but just unrelated in the case that the best of my knowledge he's never been accused by anybody in court or out of embezzling anything. It was just made up by ChatGPT's algorithm.

**[00:03:02] Eugene Volokh:** So real, I take it smells a rat, figures out that something is wrong. Actually asks ChatGPT for a portion of the complaint related to Walters. ChatGPT output something that purports to be that, realizes that this can't be right and my guess is the complaint doesn't mention, but my guess is he passes it along directly or indirectly to Walters. My sense is that they know each other. I think Walters has written for Riehl's site in the past. So, so in any event, Walters then sues OpenAI saying, "Look [laughs] you guys are libeling me."

**[00:03:36] Eugene Volokh:** And that's where the case stands. It's filed in Georgia State Court now, seems pretty likely it's going to end up in Georgia Federal Court, because it looks like the parties have different citizenship. The lawsuit is probably going to be over more than \$75,000. That's enough for federal jurisdiction. So it probably will end up being in federal court, but it's in court now.

**[00:03:57] Jeffrey Rosen:** LyriSSa tell us more about the Georgia case. And then, what, what are the legal issues? What would Walters have to prove to, prove the defamation and, and what are his chances?

**[00:04:09] LyriSSa Lidsky:** Well, one thing that I find so interesting about this case is [laughs] that first off, the, the requester of the information doesn't really appear like he was asking leading questions to, to send ChatGPT down this false avenue. And then, it even seems as if the requester was taking steps to try to correct it, or try to verify it in some way and ChatGPT didn't take the hint and that's really interesting because ChatGPT does learn as you go through a colloquy with it when you're requesting information, it can improve as you say, "No, that's not right." You know, and, and then it can go to different sources and verify them, but that certainly didn't happen here.

**[00:04:52] LyriSSa Lidsky:** The other thing that I find so interesting about this case is it invents a document that makes the defamatory allegation more credible. I mean normally it, it kind of doubles down on the, the defamation here which is really a, a unique and interesting challenge in this regard. So in an ordinary defamation case, you have to prove that the allegation is defamatory, which is pretty easy here because they alleged that he is accused of, of being

engaged in fraud. You have to prove that it identifies the plaintiff, which is again easy here, because it identified him by name, that it was published which normally that just means that it's, it's, it's made available to at least one third party.

**[00:05:38] Lyrissa Lidsky:** I think Eugene's made a good case in his new article that, that element can be met in cases like this. And then, depending on the status of who you're defaming, you usually have to prove some degree of fault in getting it wrong. If it's a public figure or a public official, you have to prove that you knew or recklessly disregarded the falsity of what you were publishing, or if it's a private figure like this guy might be a private figure involved in a, a matter of public concern, you have to prove negligence. That you behaved unreasonably in your information gathering and and verification processes.

**[00:06:17] Lyrissa Lidsky:** But of course, applying concepts of fault to a non-human entity is quite difficult and I think that's the biggest challenge of these OpenAI cases and this one and those that are likely to come.

**[00:06:30] Jeffrey Rosen:** So I hear you saying that the most challenging elements here are, first publication and Eugene has argued that simply making the publication available to one third party council's publication. And then, second the question of unreasonable behavior. Can, can a machine know or recklessly disregard a falsehood or behave unreasonably given that machines don't have intent? Eugene, do I have that right and tell us about those challenges?

**[00:06:58] Eugene Volokh:** Right. Just to be precise it's well established the communication to one person is enough to satisfy publication for libel law purposes. Classic example is a, is a false job reference. If you if you say, "Oh, this person worked for me and he stole money from me." And that's false, even if you just say it to one person that, that's libel. Let's say if you say it in writing, it's libel.

**[00:07:20] Eugene Volokh:** The question is whether this publication requirement would play out the same way when the publication is by an algorithm rather than by a human being directly and I think the answer is probably yes. But I agree entirely with Lyrissa that this is, that it's complicated to figure out how you ascribe mental state to an AI. I think the answer is that you need to ascribe mental state to the company that produces the AI. One way of thinking about it is imagine you get hit by a self-driving car and your claim is, "Well, this car was, is a defect, it was defective because it was essentially negligently designed." Because that's, that's what design defect liability under product liability law ends up being. It's negligent design.

**[00:08:04] Eugene Volokh:** Well, negligent on whose part? The creators. Right? So likewise, the claim might be this software was negligently, let's say ChatGPT was negligently designed in a way that that publishes libels and there was a more, there was a better more reasonable design, or if the claim is based on knowledge or recklessness to falsehood, it would be not what ChatGPT knows, because it's true it doesn't know anything. It's what OpenAI knows.

**[00:08:34] Eugene Volokh:** So if, for example, the plaintiff in this case had sent a letter to OpenAI or email presumably, and said, "You guys are libeling me, stop." And OpenAI didn't do anything, didn't add any code to try to block that output of that kind of information and the like, then in that case maybe at that point the, the knowledge or recklessness as to that particular false statement requirement is satisfied, not because of what ChatGPT knows, but because of what, of what the human beings at OpenAI know. That would be my sense, but there I'm much less clear- much less solid ground, because we are talking here about adaptation of old rules to fairly new technology and it's an interesting question how courts will will handle that.

**[00:09:20] Jeffrey Rosen:** Lyrissa, do you agree or disagree about the idea that it's the creator of OpenAI whose knowledge is relevant and that, that knowledge could be satisfied if they have noticed or are there different ways of looking at the question of knowledge and intent when it comes to AI?

**[00:09:36] Lyrissa Lidsky:** I think there are different ways of looking at the question of knowledge and intent. I don't think we have clear-cut answers, because our, our rules are just not designed for this scenario. They're designed for a very different scenario where a human actor is making these decisions. And so, one of the questions I, I think we have to gravel with is if you are putting a communication system out there that gets it wrong X percentage of the time, that invents hallucinations that could be false and defamatory X percentage of the time, is that act itself sufficient to base liability on?

**[00:10:18] Lyrissa Lidsky:** Is the knowledge that you're going to be systematically putting out that false information on, you know, in a certain percentage of cases enough to, to make you culpable when your product, your communications product inevitably causes harm? And I do think, Eugene, makes a good point that it's almost like a defectively designed product. That on, on policy grounds, you have to think about though, don't we need OpenAI to have some room to innovate so that it learns and improves? If we don't give any, any play in the joints for the innovation to occur, we'll never get to better, more truthful, more accurate, more reliable search results.

**[00:11:03] Jeffrey Rosen:** Eugene, there have been a number of cases that you've written about of AI hallucination, making stuff up some of which is defamatory like hallucination about my GW Law School colleague Jonathan Turley, and some of which is simply false like the brief filed in court recently that was hallucinated. Tell us about some of those well-known cases that you've written about and what different legal issues that they raise.

**[00:11:30] Eugene Volokh:** Right. So I think this is the first libel lawsuit that has been filed. There had been a threat of, of a lawsuit had been publicized. There was a mayor of a small town in Australia who sent a demand letter to to OpenAI saying, "Stop publishing this allegation against me, your own allegation nobody had made it before against me that I, that I was guilty of, of taking bribes in a bribery scandal." That was apparently hallucinated by ChatGPT.

**[00:11:57] Eugene Volokh:** Likewise, I ran some I gave it some prompts at one point, I just wanted to know what it had to say about a particular person who was accused of, of of fraud and

he said, and it said, "Oh, he he was actually, he pleaded guilty to federal wire fraud charges." All made up. I also asked, "Is there a sexual harassment problem in American law schools and give five examples?" So he gave five examples, two of which were actually examples of stories that had been in the news, and three, which again as best I could tell were entirely made up. One of which was about Jonathan Turley. And then, he ended up writing a story about that, that then led to a lot more coverage. So, so yes. It's something that ChatGPT does quite often.

**[00:12:39] Eugene Volokh:** So again, remember I asked it for five examples and two were valid and three were not, I don't have a quantifiable data, but it's roughly that. Like if you ask it about things that that are pretty specific, it seems pretty good in general if you ask for kind of a general summary of something, it often does a good job, but if, but if you ask about things that are quite specific, it will very often, very often make false statements.

**[00:13:04] Eugene Volokh:** Now as you point out, a lot of the false things that it says may not be libelous because it don't damage anyone's reputation, except they may damage the reputation of someone who relies on it. [laughs] so just yesterday there was a filing in federal court where a lawyer explained after having been berated by a judge and at the demand of the judge who said, "Explain yourself, counselor."

**[00:13:27] Eugene Volokh:** He filed that affidavit where he explained why it was that he filed this motion that had made up precedence. And he says, "Well, you know, I don't practice much in federal court, our law firm didn't seem to have access to federal court databases." He apparently didn't know of all the free online federal court resources that are out there. So I just, I heard all these things about ChatGPT and how great it was so I just gave it gave it some, some requests and it wrote up this draft and motion. And I just assumed it knew what it was talking about.

**[00:14:03] Eugene Volokh:** And then, when somebody prodded me by asking, you know, are these real cases? Well, I asked ChatGPT and gave me, gave me excerpts from the cases. I said, "Is this a real case?" ChatGPT said, "Yes. It's a real case." so the problem is, actually, the problem is twofold. The problem is much of the time ChatGPT actually provides shockingly accurate results that there are news stories, which I have no reason to doubt, that the report that it performed at the 90th percentile of, of bar takers the bar exam, you know, that, that's pretty good.

**[00:14:41] Eugene Volokh:** But that actually makes it more dangerous when it makes up stuff. If it were just known to be a bunch of monkeys and a bunch of typewriters or a Ouija board or something like that, it wouldn't be dangerous, it wouldn't be useful though. It wouldn't get \$13 billion in an investment as it apparently recently got. But at least it wouldn't be that dangerous, but here you have a combination of something that is, that is often quite reliable, people end up trusting it based on experience, but that trust is often misplaced.

**[00:15:13] Eugene Volokh:** By the way, note everything I say applies equally to human intelligence, right? If AI has just invented, has just proven it's available of make, to make stuff up and and just blatantly make make false statements without any shame and with, with maximum seeming confidence, that's just like people.

**[00:15:35] Jeffrey Rosen:** Yeah. Indeed. And I want to ask about punishable and unpunishable falsehoods in a moment, but just a few more beats on defamation. Lyrissa, not everyone agrees with Eugene that liability for the chatbot is appropriate under current definition law. Ronnell Andersen Jones from the University of Utah has said that if the defamatory faucet is generated by the AI chatbot it's hard to con- to conceptualize this within our defamation law framework, which presupposes an entity with a state of mind on the other end of the communication and she says that some scholars suggest that a remedy here resides more on a product liability model than a defamation model. U, what, what do you make of that argument?

**[00:16:17] Lyrissa Lidsky:** Well, I think Eugene has considered that argument and as well as another author Nina Brown has considered that argument. And I think that's somewhat where we're going to have to go, because if you were putting out a defective communications product that's producing these erroneous results that cause harm a, a substan- you know, a substantial percentage of the time, it seems fair to make that industry internalize some of the costs of the harm that it's, it's, you know, imposing upon society.

**[00:16:49] Lyrissa Lidsky:** And so I, I think Ronnell is correct that the culpability that traditional defamation law looks at doesn't quite fit here and that's why we need to look to other areas of tort to say what is the culpability? And the culpability is putting out a communications tool that's, that's basically defective. Now the hope is it's defective now but it's going to get better and better over time and how do we develop a liability regime that allows it to improve and get better and better over time so that it is accurate.

**[00:17:22] Lyrissa Lidsky:** I think one thing that the story about the, the legal case involving the lawyer relying on ChatGPT illustrates is that if you're a journalist looking to ChatGPT as a reliable source without verification, that could very well constitute actual malice. And if you publish based on a ChatGPT result you could definitely be liable for defamation and I don't think you'll be able to easily defend saying, "Well, I published it because ChatGPT told me to."

**[00:17:53] Jeffrey Rosen:** Eugene tell us more about your solution. You have argued that, that when the actual malice standard applies, the standard might be satisfied if the AI company gets actual notice a particular spurious information and in practice this would require companies to implement a notice and blocking system similar to notice and takedown under copyright and trademark infringement law. Tell us about how that would work in Miss Walters case and, and elsewhere?

**[00:18:19] Eugene Volokh:** Sure. So I think there are basically two ways in which the AI companies could be held liable under existing law. Existing law as adapted in modest ways to this new technology. One is when the so-called actual malice standard applies. Of course, actual malice is this weird legal phrase that doesn't actually mean malice. It means knowledge that statement is false or is likely false, also known as knowledge or recklessness or knowledge or reckless disregard.

**[00:18:48] Eugene Volokh:** So in the actual malice standard applies there has to be actual knowledge on the part of the company. Let's say on the part of OpenAI as the ChatGPT. And

that would usually be satisfied by a takedown demand. That just like let's say you're a reporter, you're writing about a public figure. The standard there is knowledge or recklessness standard and you thought that a certain statement that you keep mentioning in your columns is correct. Okay. Then, you're off the hook because you as to those columns, because you sincerely thought that statement was correct, but then the, the public official emails you and says, "No. No. This is false and probably false, and here's all the evidence."

**[00:19:28] Eugene Volokh:** At that point, if you keep, if you keep publishing that statement going forward, then in that case, you would be liable. So so that would be, again, this kind of notice and blocking regime where once they're unnoticed they have to create some code and I don't think it's, it's that difficult to create, just maybe some post-processing code that takes the output of say ChatGPT and identifies names and looks it up on a, looks them up on a list of, of notices that it has gotten. That particular statements about particular people maybe false and just stops outputting those statements. It requires some coding but, you know, OpenAI is in the coding business.

**[00:20:07] Eugene Volokh:** The other theory is a negligence-based theory. Now when does libel law allow negligence-based liability? One possibility is as the statements matters of purely private concern. So let's say some allegation that somebody, especially somebody relatively low profile had had an affair let's say or something along those lines. Another possibility is if it's a statement on a matter of public concern about a private figure and there's proven actual damages.

**[00:20:36] Eugene Volokh:** So if for example somebody lost the job or lost a business opportunity or even lost friends perhaps, lost social opportunities as a result of a statement. That, that may be very hard to prove in many situations but some- sometimes they could say, "Yeah. You know, somebody stop doing business with me and when I asked why, he said, 'Because I looked you up and it accused you of being a child molester.'" He says, "No. No. This is all made up." That's have actual damages and even if the accusation of coming, committed a crime is a matter of speech on a matter of public concern, I'm a private figure I should be able to prevail in a negligence theory.

**[00:21:09] Eugene Volokh:** So what does negligence look like? I think it's not a question of do you use negligence liability or do you use products liability? I think products liability offers a helpful analogy for negligence law. Technically to be pedantic, products liability law does not apply to for reputational damage. If you look up, for example, a very important work on products liability or very influential, they're a statement of torts on product liability. It specifically has a provision that says that this only applies to situations where product defects cause damage, physical damage to person or property, but so pure economic loss, lost business opportunities, you can't sue for products liability based on that, but you can sue for negligence and products liability law has a lot to say about what it means to be negligent when you are deploying a product.

**[00:22:05] Eugene Volokh:** So one way you can negligent, you can have product liability is by engaging in negligent design, in designing a product that is unreasonably dangerous, which is to say that there are alternative, reasonable alternative designs that would have avoided the danger.

So I think in this kind of negligence-based libel lawsuit, when that's of, when negligence is available as a theory under liable law, you can import these concepts from product design law and say, "Wait a minute, this is a badly designed product. It hallucinates more often than a better design product would or, or it reports things between quotation marks that it has no business reporting between quotation marks because a reasonable designer would recognize that the stuff between those quotation marks may not actually be literally accurate." so that, those I think are the main theories for libel by AI.

**[00:22:59] Jeffrey Rosen:** Lyrissa, what you make of Eugene's proposals and how would they work in practice? Would, would Walters object that the AI had hallucinated about him and the machine would be programmed not to make future statements about him? And if someone felt that AI were unreliable and accurate on a systematic basis, could legislature simply ban its use across a whole range of information gathering purposes consistently with the First Amendment or not?

**[00:23:26] Lyrissa Lidsky:** Okay. [laughs] There's a lot in that question so let me, let me take it systematically. So one of the things that Eugene talked about is adapting negligence concepts by analogy to products liability law, which I think is a good and valuable suggestion. Who would do that? That would most likely be done by a judge interpreting tort law, the common law and reasoning by analogy and it, you know, it's a stretch. It's a stretch of existing law. It's not clearly covered under existing law concepts of negligence so you're going to have to stretch and reason by analogy and have some common law creativity there and have a judge willing to engage in that kind of creativity.

**[00:24:08] Lyrissa Lidsky:** A notice and takedown liability regime, probably that would most likely be enacted through legislative work. So you're going to have to have a, a law saying if you have... I mean Eugene's arguing you could do it by as an analogy on the actual malice point, but I think the easiest and most least complicated way to get there would be through legislative work.

**[00:24:32] Lyrissa Lidsky:** And then, the last question you asked me was about constitutional law principles and is any of this even really fully protected speech? And I think the answer is mostly the answer is it is protected speech because we need to protect it in order to, to protect our right to receive information. It's a valuable source of information and we worry about the effects of, of saying that it was not protected speech, it, you know, does open up to governments to just entirely shut down access to whole media of, of new communication.

**[00:25:06] Lyrissa Lidsky:** I think some of the TikTok fans are, are instructive in this regard in terms of government motives but, but it's obviously a complex adaptation problem that needs to proceed on multiple different fronts at, at once, by people who have a sophisticated understanding of how large language models work. And that's where I think it gets really difficult.

**[00:25:29] Jeffrey Rosen:** Absolutely. And what makes it so fascinating. Eugene many theories of the First Amendment emphasize the self-expression and reason of the speaker as Justice



Brandeis said in the Whitney case, "Those who won our revolution believe the final end of the state was to make men free to develop their faculties."

**[00:25:47] Jeffrey Rosen:** And that interest obviously does not apply to AI, which has no faculties of reason, but Lyrrisa mentioned the listeners interest in information gathering, and indeed the law does protect forms of speech like commercial speech which are not forms of self-expression but contribute to the listeners interest in information gathering given that under current First Amendment doctrine, could AI speech be regulated like commercial speech about the, may- maybe more intensely than purely political speech, but not banned altogether, or do we need to rethink our entire theoretical framework for First Amendment regulation?

**[00:26:23] Eugene Volokh:** It's an interesting and difficult question. My tentative thinking is that AI output should be as protected as other speech. And one way of thinking about it is imagine that the government were to say, "You know, Aristotle, he had all these weird thoughts, they are kind of obsolete now, people maybe take them too far. Some of them are sort of dangerous, not our kind of guy. We're just gonna, we're just gonna ban republication of Aristotle."

**[00:26:56] Eugene Volokh:** Does that violate Aristotle's rights? [laughs] I doubt it. You know, Aristotle I think is beyond the reach of human law at this point. Now I suppose you could say, "Well, what about the publishers? People might want to republish it so they have the right, even though, even though they didn't create a word of it, they didn't even create a trans- the translation. They have their own free speech right."

**[00:27:17] Eugene Volokh:** You could say that though equally then you could say, "OpenAI, the human beings run OpenAI equally have the the right under the same rationale to, to put out ChatGPT and have its output be protected. True they're not the ones who actually generate the specific details of the output, but neither does the publisher of Aristotle as neither is he the author of of what Aristotle wrote."

**[00:27:41] Eugene Volokh:** I don't think that's how we deal with the Aristotle ban. I think we'd say, you know, the rest of us have, have a right to read Aristotle and, you know, some of the stuff may be wrong, but it'll be up to us to figure that out. And to the extent that some of it may be, there are some things published by by people who are dead might be libelous. Well, you could imagine some liability against publishers for that libel, but again that would be constrained by First Amendment principles.

**[00:28:11] Eugene Volokh:** I'd say the same with regard to with regard to ChatGPT. You don't have to have some rights holder with self-expression rights present to have a First Amendment claim there. Just like you don't have to have Aristotle's First Amendment rights be in play with regard to the Aristotle ban. It's enough if restriction on ChatGPT interferes with the rights of the public or rights of users to use it to read its output.

**[00:28:40] Eugene Volokh:** Another way of thinking about it is in recent years many lower courts have concluded that the First Amendment includes a right to gather information cheaply. This has come up in situations where there are restrictions on video or audio recording. And the Supreme Court has never quite squarely confronted that, but it has certainly talked about First Amendment protecting the rights to gather information. Well, that too is an important listener interest. Many people do use ChatGPT to gather information, although I hope they use it with suitable skepticism as much information gathering should include.

**[00:29:15] Eugene Volokh:** So I think one way or the other, I think we get to full First Amendment protection, whether we say OpenAI has rights over software that, that it did create and it did program in certain ways. OpenAI is deliberately trying to to include certain messages in some respects, exclude other messages. So we could say it's OpenAI's First Amendment rights that are at stake, because of their self-expression indirectly through ChatGPT, or we could say it's the rights of readers as readers or as information gatherers they're at stake. But one way or the other, I think the First Amendment is in play. Of course, there are exceptions to First Amendment and one of them is liable.

**[00:29:53] Eugene Volokh:** So that doesn't preclude libel liability, but that suggests that the First Amendment defenses that have been erected by the court in libel cases should apply, I tentatively think to AI output as well.

**[00:30:07] Jeffrey Rosen:** Lyrissa, you raised the question of broader protection under current law for lies, unpunishable lies, and, and, and they're broad. The Supreme Court has said that some lies are constitutionally protected even if the speaker knows the statements are false. Those include false statements about philosophy, religion, history, the social sciences of the arts and the like.

**[00:30:28] Jeffrey Rosen:** That's from a United States versus Alvarez where five justices emphasized the danger of putting the state in the business of deciding what's an opinion and what's a fact, and therefore allowing broad protection for lies when they're not defamatory. How would that apply to the many other cases of AI hallucinations that are, are lies that may not be defamatory such as the made-up brief and so forth? And more broadly, how should the law think about non-defamatory lies generated by AI?

**[00:31:04] Lyrissa Lidsky:** So US versus Alvarez was a Supreme Court case that involved someone who falsely claimed he had won a Congressional Medal of Honor and he was prosecuted for a crime under the Stolen Valor Act that in that law was a federal law making it a crime to falsely claim you won these, these medals. And the argument was that we needed to punish people who did that because it would undermine the value of the medal to those who had legitimately won them.

**[00:31:33] Lyrissa Lidsky:** And the Supreme Court struck down that law and they did it in a very interesting way. So first off, they didn't say that lies are particularly valuable. They did find the lie in this case to be protected, but it's because the lie didn't cause what I would describe as a cognizable harm. Now there's certain kinds of harms that trump First Amendment protection like

defamation. A truly defamatory statement can trump any First Amendment protection that a defamatory lie could have. So the defamatory part trumps the lie part in terms of that. But I think you have to understand Alvarez in light of this strong strand of distrust of government in our First Amendment jurisprudence, that we don't want the government set up as an arbiter of lies particularly when those lies don't cause serious kinds of harms that we are concerned about giving compensation for.

**[00:32:38] Lyriisa Lidsky:** And so, the court said, "Don't ask whether the speech is valuable first, ask first whether it's harmful." And they found that it wasn't harmful in that case because in 30 seconds you could Google and find out that the guy wasn't really a Medal of Honor winner. So it relies on trusting us as citizens to behave rationally, to use our critical faculties to sort truth from falsity and a distrust of the government to be the one that's telling us what's truth and, and what's false, particularly when the speech doesn't appear to be especially harmful.

**[00:33:12] Lyriisa Lidsky:** But defamatory falsehoods have always been punishable under the First Amendment if they are, you know, particularly even about public officials and, and celebrities, if their intentional lies, knowing falsehoods or their reckless, recklessly disregarded falsity types of statements.

**[00:33:34] Jeffrey Rosen:** Eugene, given Alvarez, are there other forms of non-defamatory lies generated by AI hallucinations that are cognizable under current law? And, and we've talked about a range of them from the made-up briefs, to the simply false historical information. Does the current law give that kind of lie a broad breathing room and, and should it?

**[00:34:01] Eugene Volokh:** Right. So I think the analysis again here would be the same as for false statements put out by other by other entities, which is a weird way of saying put out by humans. [laughs] We are now the other entities. [laughs]

**[00:34:13] Eugene Volokh:** So, so let me give you an example. There is the false light tort, which generally deals with false statements that are offensive, not because they damage reputation, but because it's just offensive to have certain false things said about you. So for example, if somebody falsely says that I have cancer, then, you know, it's not that it's bad for my reputation, it's that it's just offensive for people to, to be, for have that be said about them, maybe have them be the objects of pity, that they think or even of sympathy that they think is misplaced because like people are feeling bad for them and people are, are expressing kind of their, their, their support when there's really no, no reason to do that.

**[00:34:54] Eugene Volokh:** So imagine that AI that a ChatGPT just hallucinates that. You could imagine a false like claim. Note how this ties into Lyriisa's point that, that, that false light tort is generally seen as constitutionally valid, because it involves a, a particular fair- fairly individualized harm. It's actually a harm to a particular person, not as severe probably as harm of damage reputation, but severe enough that the Supreme Court in a couple of cases has allowed that kind of liability for false light, and probably Alvarez doesn't undo that.

**[00:35:26] Eugene Volokh:** Here's another example. Let's say that ChatGPT outputs certain statements that are that like certain advice about how to treat some illness. And that advice is just false and that causes physical harm. Generally speaking, even negligently conveyed information that causes physical harm might in theory be subject to liability. And certainly knowingly false information that causes physical harm might lead to some liability.

**[00:36:00] Eugene Volokh:** It's not open and shut. There's actually a Ninth Circuit case, there's a federal appellate case from many years ago where there was a lawsuit against a mushroom encyclopedia. The publishers of mushroom encyclopedia which report, which in the encyclopedia describes some mushroom as safe to eat even though it was deadly. You'd think there might be liability there. No, says the court. No liabilities or at least on a, on a negligence theory.

**[00:36:24] Eugene Volokh:** So there's some uncertainty there, but at least there's some possibility of that kind of liability. Now to be sure maybe ChatGPT or OpenAI could deal with some of that through our suitable disclaimer by requiring people to click on something say I acknowledge that this may not be accurate and I wave any right to sue OpenAI. So maybe that kind of disclaimer might be upheld although that's not clear especially when physical harm is involved, but also you could imagine this causing harm to a third party.

**[00:36:54] Eugene Volokh:** Let's say, for instance, that that ChatGPT tells me this mushroom is safe to eat, even though it's been actually specifically been alerted that, that is false and that's a hallucination that ChatGPT routinely outputs. And then, I feed the mushroom to my wife. She didn't wave her rights. So as a result, she could in principle sue she could sue me [laughs] for foolishly relying on ChatGPT, but she could in principle sue OpenAI for for wrongfully passing along that information. So in theory, there could be such liability.

**[00:37:26] Eugene Volokh:** Here's what I don't think there would be liability for. Imagine it, ChatGPT says, "Yes. 9/11 was an inside job that it was the American government that destroyed the, the, the Twin Towers." That's false, but it's, unless there's some particular person it's accusing that, that's something that is constitutionally protected from liability, even if it's a knowing falsehood under case, under Alvarez and other such cases.

**[00:37:53] Jeffrey Rosen:** Very helpful to have that distinction between falsehoods directed in particular individuals and those that are more generalized. Lyrissa, The Wall Street Journal recently reported about a chatbot that impersonated Taylor Swift available through Banter AI and invited people to ask about her music or her relationships. And of course, the use of AI to impersonate political candidates and celebrities may well be rampant. To what degree does current law make that actionable and how does law have to adjust?

**[00:38:25] Lyrissa Lidsky:** That's a great question. Deepfake's used to create a defamatory assertion effect can easily fall into a defamation claim but I, I don't, with Deepfake, I think we're concerned beyond just defamation. We're concerned with disinformation, military disinformation. There's already been some uses of that in the Ukraine context is creating false videos to suggest something that never happened.

**[00:38:51] Lyrissa Lidsky:** I think as, as you pointed out the fraud context is one, in, in which Deepfakes are going to be a serious concern and there's no problem with punishing fraud. It's, it's, you know, creating a Deepfake to perpetrate a fraud is just going to be considered a form of conduct unprotected by the First Amendment. Disinformation, as Eugene pointed out is a harder question because we normally distrust the government when it goes about policing disinformation and setting itself up as the Arbiter of Truth, unless we have a specific cognizable harm to an individual that we can, we can point to.

**[00:39:27] Lyrissa Lidsky:** And so Deepfakes used to perpetuate disinformation are unlikely to be capable of being regulated specifically by, by the government although I can envision some transparency requirements. And I think in general with ChatGPT and with some of these Deepfakes is some of our solution to these problems are going to be in mandating transparency around the issue that they are fake.

**[00:39:53] Jeffrey Rosen:** Eugene, tell us about other legal issues that could arise out of the misuse of name and likeness and some open legal issues here.

**[00:40:03] Eugene Volokh:** Yeah. So libel is one kind of tort that is based on what people communicate false light is another. There is the so-called right of publicity which is the right to exclusive right to use one's name or likeness for certain kinds of commercial purposes, but that is not, in many respects considerably narrower right after all, for example, an unauthorized biography of a person uses their name and likeness and life story for commercial purposes.

**[00:40:33] Eugene Volokh:** And newspapers write all the time for commercial purposes hoping to make money, you know, maybe, sometimes they don't succeed but they try at least by writing about people likewise you could have fictional works that that use names and likeness of likeness is a famous people, either living or recently dead. Forrest Gump is a classic example, Midnight in Paris there are lots of others.

**[00:40:55] Eugene Volokh:** So generally speaking, if all that that OpenAI, or let's say all the ChatGPT does is output some item that mentions Taylor Swift or let's say somebody says write me a short story about Taylor Swift getting together with Elvis Presley to write a song together. And it says, "Oh, here's a short story about it, it's about..." It mentions their name and likeness, it's for commercial purposes in the sense that OpenAI is a, is a business corporation and ChatGPT for at least is a subscription-based product for which people pay money, but that's not going to be seen I think as a right of publicity infringement precisely because it is that kind of commercial purposes still a fully protected purpose of providing entertainment, providing fiction or providing sometimes factual reporting.

**[00:41:47] Eugene Volokh:** Something that, that sometimes comes up when we talk, not about ChatGPT, but about visual AI programs like Dolly and, and I think Stable Diffusion and various others is what happens if somebody creates, uses them to create a picture of a person, of a real person, and some of them are, may refuse to do that as a matter of design of their own design decisions, but let's say they do create a picture of a real person. And there, it's a little bit more complicated.

**[00:42:18] Eugene Volokh:** So for example, if the picture is pornographic there's a good deal of case law for the proposition that in fact the right of publicity extends to at least commercial pornography and maybe the related tort of misappropriation of name or likeness may extend even to non-commercial pornography. On the other hand, if it's a political parody. Let's say, or even non-political parody, pretty clearly going to be protected by the First Amendment. And then, if it's just a straight up, I want a picture of Taylor Swift to hang on my wall. Then that might in fact be a right of publicity infringement and possibly also a copyright infringement if it copies a particular photograph.

**[00:42:58] Jeffrey Rosen:** It's so fascinating to think through with you the effort to apply existing common law torts to this AI universe just as Louis Brandeis did in the face of new technologies and rethinking privacy at the turn of the 20th Century. Let, let's pull back to the broader ways that AI will transform our free speech universe in a world where history is rewritten and when the information people find about the past online is generated by AI and can no longer be trusted as entirely reliable, how will that change our sense of truth, falsehood, and, and more generally of reality?

**[00:43:39] Lyrrisa Lidsky:** Well, I, I do think it's a, it's a threat to our our concept of truth and our insistence that public discourse have some anchor in truth. I think that, that anchor perhaps has already been unmoored somewhat by the internet, social media just the information overload that we're facing now and this will further destabilize people's sense of where they can go for reliable information.

**[00:44:07] Lyrrisa Lidsky:** I mean I don't believe in apocalyptic theories. I'm, I'm a, I'm an optimist. I think it relies more heavily on us exercising our critical thinking faculties, being careful about where we go for information, being careful about who we learn to rely on as critical sources of information and it could be that we are going to have to apply a great deal of skepticism to anything that comes out of OpenAI. I find OpenAI valuable as an expert looking to verify things that I already know quite a bit about.

**[00:44:40] Lyrrisa Lidsky:** I think for somebody starting research not knowing anything about a topic, it's a very dangerous tool because you won't be aware of what it's giving to you that's utterly false. And so, and perhaps I'm overestimating my own abilities to discern what's utterly false in the results it's giving me even in the area where I have some degree of expertise.

**[00:45:03] Jeffrey Rosen:** Powerfully said. Eugene, time for final thoughts and what I hope will be the first of many discussions with both of you because we all have so much to learn about this brave new world. What are your thoughts about the degree to which AI challenges the founders' confidence that as Brandeis put it, "Freedom to think as you will and speak as you think are means indispensable to the discovery and spread of political truth?" And more broadly, give, give us a, a road map to the broad and important legal and constitutional issues that will be presented by AI in the years to come.

**[00:45:39] Eugene Volokh:** Well, the framers of the Constitution were no strangers to falsehoods or to statements that are maybe even not literally false but misleading or the opinions

that are unfounded. They protected the freedom of speech and of the press despite the, a known reality that it could be, that they could be used in dangers and irresponsible ways. And at the same time they recognize there are going to be exceptions to that such as libel is the clearest clear such example.

**[00:46:10] Eugene Volokh:** So the, theory I think has never been that, that freedom of speech is perfect at uncovering the truth. It's just better than the alternative. It's better than non-freedom. But this having been said it may very well be that the new technologies do make it more difficult for people to, to figure out what's really true and require, require them to maybe learn some new habits or maybe relearn old habits, right? So, so it used to be I think that, that people generally understood that there were more trusted sources and less trusted sources, and usually the sources that were broadly available tended to be ones that were someone more trusted and somewhat more trustworthy, although far from perfectly trustworthy.

**[00:46:57] Eugene Volokh:** Newspapers television broadcasts and such. Then, I think with the growth of the internet it became a lot easier to access all sorts of things. That somebody could just do a Google search and if Google happened to point to some particular site they could, they could kind of assume especially in those, things that are low stakes for them that the material there is right. I think, you know, maybe I'm, maybe it's mistaken but it's not really worth my time to investigate further.

**[00:47:22] Eugene Volokh:** On the other hand, if it ends up being that the great bulk of all of the pages out there are actually created by by AI products that that are being used as part of disinformation campaigns the, the, the potential for error may become high enough that people may need to relearn this and say, "Well, wait a minute. Where's this from? Is there something that indicates it's from a, a mainstream publisher?" For all the flaws that mainstream publishers have or is this maybe just a, a something that pretends to be a mainstream publisher or pretends to be authoritative even though it's not authoritative.

**[00:48:03] Eugene Volokh:** And by the way, this is something that people have had to deal with regard to forgeries all the time in the past, right? That people, my understanding, at least, I've heard it said and maybe it's all made up. But I've heard it said that back when people were generally when there was a lot less writing, there, there was no print, people sort of assumed a written document, must be real, it's written down. At least many people did and of course there were lots of forgeries that exploited that as there still are.

**[00:48:28] Eugene Volokh:** So people might people I think over time became more skeptical and may become, have to become more skeptical still. They need to maybe use special browsers that actually indicate where, where really this site is hosted let, let's say and whether it is, whether it has some validation mark. Of course, the other thing is the media would have to re-earn our trust that I think part of the things that we've seen is media become somewhat less trustworthy. And it would be good if they tried harder to, to make sure that, that what they are saying is accurate and to the extent possible objective.

**[00:49:04] Jeffrey Rosen:** LyriSSa last word to you in this really important discussion. Do you agree or disagree with Eugene that's in the world of AI, the responsibility primarily should remain on citizens to distinguish between true and false information with, with notable exceptions like defamation law? And what do you see is the thresholds for the First Amendment of the Constitution and AI in the years to come?

**[00:49:28] LyriSSa Lidsky:** Well, I do believe along with Eugene that the responsibility primarily rests on citizens and the First Amendment has long assumed that citizens are rational actors and capable of discerning truth and falsity and only when there are system failures that, that don't permit them to do that can we allow things like transparency requirements, disclaimer requirements, disclosure requirements. And we may discover that some of those are necessary to assist us in discerning truth and falsity and there is some, some indication in First Amendment precedent that those are sometimes allowed but they're all almost always preferable to complete bans on access to new media of communication.

**[00:50:11] LyriSSa Lidsky:** So for example, the, the TikTok bans or, or any attempt to ban use of ChatGPT is of, is of great concern if, if the government endeavors to do that. So I do think one, the our First Amendment theory has long depended on us as citizens to rise to our best capacities because we don't trust the alternative. And I think that's going to be continued, continue to be true regardless of what generative AI does.

**[00:50:42] Jeffrey Rosen:** Beautifully said. First Amendment Theory depends on us as citizens to rise to our best capacities, because the alternative is worse. Thank you so much, Eugene Volokh and LyriSSa Lidsky for a just a, a centrally important discussion about free speech AI and the First Amendment as I said I hope it's the first of many. Thank you both.

**[00:51:04] Eugene Volokh:** Thanks very much for inviting me.

**[00:51:05] LyriSSa Lidsky:** Thank you.

**[00:51:10] Jeffrey Rosen:** This episode was produced by Lana Ulrich, Bill Pollock and Sam Desai, was engineered by Greg Scheckler. Research was provided by Rosemary Li, Sam Desai and Lana Ulrich using their human intelligence and capacities of reason. Please recommend the show to friends, colleagues or anyone anywhere who's eager for a weekly dose of constitutional illumination and debate. Sign up for the newsletter at [constitutioncenter.org/connect](https://constitutioncenter.org/connect).

**[00:51:35] Jeffrey Rosen:** And always remember that the National Constitution Center is a private non-profit. We rely on the generosity, the passion, the engagement, the intelligence of people from across the country who are inspired by our non-partisan mission of constitutional education and debate. Support the mission by becoming a member at [constitutioncenter.org/membership](https://constitutioncenter.org/membership) or give a donation of any amount to support our work including this podcast at [constitutioncenter.org/donate](https://constitutioncenter.org/donate). On behalf of the National Constitution Center, I'm Jeffrey Rosen.



