SCRIPT - TECHNICALLY OPTIMISTIC - SEASON 2 - EP. 1

RAFFI: We live inside of a giant pile of data.

[SFX: A car beeps to unlock, a door closes, someone gets in**]**

RAFFI: Digital life runs on information. It *is* information.

[SFX: RAFFI: "Hey Siri, can you give me directions to the grocery store?"
SIRI: "It adds ten minutes to your route."]

RAFFI: But not just digital life. All kinds of things in the physical world are the way they are because of data.

[SFX: Car beeps.
GPS: "Arrived."**]**

RAFFI: Traffic patterns. Train schedules.

[SFX: Car door opens, footsteps**]**

RAFFI: The price of eggs.

[SFX: The electronic noise of an automatic door; din of a store**]**

RAFFI: Look. We all have a complicated relationship with data. It's given us personalization. It gives us insights into how the world works.

[SFX: A cash register ringing items up**]**

RAFFI: It lets us be smarter in how we interact with each other.

[SFX: CLERK: "Alright. It comes to eight sixty-four."
RAFFI: "Okay. Cool."
A credit card machine dings indicating a successful transaction.**]**

RAFFI: But…and I know this isn't news…

[SFX: CLERK: "Alright, you have a good day."
RAFFI: "Thanks a lot. I appreciate it."
Footsteps; electronic door again**]**

RAFFI: Data is being collected from us. Constantly. From you, from me, from people everywhere.

[SFX: Person getting back in car.
SIRI: “Where would you like to go?”
RAFFI: “Ok, now let’s drive to work.”
SIRI: “Getting driving directions to work.
A car accelerates.**]**

RAFFI: We share personal information all the time. Tons of it we do deliberately, on purpose. We post to social media, we send emails, we send text messages.

[SFX: Din of driving.
RAFFI: “Hey Siri, can you send a message to Erikk; tell him I’m running late?”
SIRI: “Your message to Erikk says…”**]**

RAFFI: And most of the time, we think of these things as private.

[SFX: SIRI: “Do you want to send it?”**]**

RAFFI: We might not think about the tech giants who own the platforms we’re using to communicate.

[SFX: Din of driving, turning the radio on.
NEWS CLIP (*NPR*): *“Facebook’s share price tumbled today. The tech giant is on the defensive…”*
Channel change on the radio.
NEWS CLIP (*NPR*): *“Amazon is now storing government data on the cloud, selling internet-connected door locks, making the popular Alexa smart speaker…”***]**

RAFFI: Or, even if we know it’s not private, we think of our data as protected, safe. This doesn’t always work out.

[SFX: Radio dial changing stations
NEWS CLIP (*NPR*): *“A massive hack of the company Equifax has compromised the personal information of as many as a hundred forty-three million Americans…”*
Channel change on the radio.
NEWS CLIP (*NPR*): “*There’s been a massive data breach that’s compromised sensitive information for millions of people. This one happened at Capital One…”***]**

RAFFI: And then there's all this other information we shed, almost passively, involuntarily, without even knowing it. And not just our phones, our smart TVs or kitchen appliances, or robot vacuums, or exercise machines…

> **CLIPS:**
> **[Satya Nadella]:** "*If you look at our lives today, you have computing everywhere. It's ubiquitous."*
> **[Tim Cook]:** *"I think most people are not aware of who is tracking them, how much they're being tracked…"*
> **[Satya Nadella]:** "*When you have computing everywhere, and all that computing is connected, you collect a lot of data. . ."*

RAFFI: And again, we know all this. We know our data is being collected. We even know that it's being used to show us ads or "targeted content".

It's a truth about life online today that many of us take for granted. But in doing so, we all might be getting a little bit…complacent.

The question is: Where does convenience end? And where does surveillance begin?

> **[SCORE CHANGE]**

RAFFI: This season on Technically Optimistic, we're taking a deep dive into your data. Who's collecting it? Where is it all going? And how is data shaping the world around us?

How is data collection changing health care?

> **CLIP [Melanie Fontes Rainer]:** *"What does it mean to now be targeted because of who you are, and have your data targeted because of the kind of health care you're seeking and where you live?"*

RAFFI: Our criminal justice system?

> **CLIP [Kashmir Hill]:** *"Should all of the people whose faces are in the database be part of a lineup every time a crime is committed?"*

RAFFI: Our children?

> **CLIP [Tiera Tanksley]:** *"The kids notice immediately that there are patterns. And then they're like, 'How? How does this happen?'"*

RAFFI: And how we connect with each other?

CLIP [Ethan Zuckerman]: ***"****I am trying to develop a new model of social media based around much, much more user control."*

RAFFI: How should you think about the future of data?

At Technically Optimistic, we believe that we *can* have a data-driven world, with all the benefits. But it's *your* data. So we want to make sure that you have a say in how your data is used.

The reality is, you have the power to shape that future. And that starts with understanding how this all works.

But, we should start now. So let's do this.

I'm Raffi Krikorian. And from Emerson Collective, this is season two of Technically Optimistic.

[MUSIC]

RAFFI: In October 2023, there was a hearing in the US House of Representatives, organized by the subcommittee on Innovation, Data, and Commerce.

CLIP: [US Representative Gus Bilirakis] *Good morning, everyone…*

RAFFI: It was one of a bunch of hearings on Capitol Hill around this time focused on AI.

CLIP: [US Representative Gus Bilirakis] *Central to these discussions has always been the need for America to lead in the development of standards and deployment and what AI means for our data.*

RAFFI: Witnesses from the tech world were being called upon to share their expertise — and their opinions — about how the US government might regulate this new technology.

And for this one, they actually invited me.

CLIP [US Representative Gus Bilirakis]: *Our first witness is Raffi Krikorian. You're recognized for five minutes.*

CLIP [Raffi]: *Thank you.*

RAFFI: And I'm bringing this hearing up, not because I got to speak…

CLIP [Raffi]: *My name is Raffi Krikorian, I'm the chief Technology Officer at Emerson Collective.*

RAFFI: …but because what happened in that room shares something in common with this season of our show.

> **CLIP [Raffi]:** *We live in an age of rapidly increasing digital surveillance. And very few users understand the tradeoffs they make when they're using their phones or the web…*

RAFFI: See, our first season was all about artificial intelligence. And as we asked expert after expert what their best visions were for AI, they kept telling us: you can't talk about AI without talking about data.

That's because, on a fundamental level, AI needs a massive amount of data in order to be trained, and really in order to function at all.

So it made sense that data privacy would be very much on the agenda at this October hearing. And the panelist sitting to my left brought up the topic right away.

> **CLIP [Amba Kak]:** *"My name is Amba Kak, and I am the executive director of the AI Now Institute, and I have over a decade of experience in global technology policy. I want to make one overarching point in today's testimony, and that is that we already have many of the regulatory tools we need to govern AI systems. Data privacy law is a core mechanism that can help mitigate both the privacy but also the competition implications of large-scale AI. Data privacy regulation IS AI regulation."*

RAFFI: The US? We have no federal data privacy law on the books. Some states, like California and Illinois, they have their own statutes. But the closest Congress ever came was in 2022, when the American Data Privacy and Protection Act, or the ADPPA, passed through this committee — almost unanimously. But it stalled in the Senate. And, right now, it seems dead.

So it was a bold move for Amba Kak to come out of the gate suggesting that even though we don't have a federal privacy law, we somehow have what we need to regulate AI.

So I caught up with her a couple months later to ask her about this move.

AMBA KAK: So the move that I made in that testimony, when I was like, 'we already have all of the tools,' was a very specific one, and one that we've sort of been making more and more in an environment of almost, like, hopelessness and awe at these kind of AI systems. It's like, oh, the whole landscape has changed.

And so the move is really to remind people that we have these sort of legacy institutional frameworks globally like data privacy law, which, do they need to be strengthened? Absolutely. But like are the bones there? I would say yes. And essentially, if you stop looking at AI systems as these composite objects that have nothing to do with our existing information ecosystems, that's not really what they are. They are essentially powered by data and data is a key input, it's also an output. And therefore data privacy law becomes a really important lever across that life cycle.

RAFFI: As Amba made clear later in her testimony, the adverse data privacy effects of AI aren't hypothetical. They're here now.

> **CLIP: [Amba to Rep. Pallone]** *"The first is an obvious one, it's privacy. We are seeing new privacy threats emerge from AI systems. We absolutely do know what harms they're already causing. They're leaking personal information, they could potentially be leaking patient data in health care contexts. These privacy risks are not abstract, even if the technologies are portrayed as these abstract, magical systems. The harms are very, very real."*

RAFFI: So it's clear we need to do something to protect data privacy right away. But where do we start? What's our best first move, if we have to act fast?

AMBA KAK: Yeah, I think we've seen this solution in privacy law already. Privacy law, when it began, it was kind of a much more notice and consent regime. And so a lot of the focus was on choice and individual choice and empowering individual choice because at the heart of it, privacy is rooted back to individual autonomy, and that's a very powerful value. But I think what became very evident soon enough is that the way in which the power asymmetry in our kind of digital environment is so immense and that power asymmetry is at the root of it, an information asymmetry, right? There needs to be certain kind of bright lines around data use, so, irrespective of what the individual chooses, there should be certain limits to the kinds of data that companies collect and how long they store it for and all of that.

RAFFI: Sure.

AMBA KAK: And so in the age of kind of AI — and we didn't start this age in 2023 — it's been kind of growing in momentum over the last few years, what we do know is that the kind of incentives for invasive and more and more longer term data retention of data it already existed, but those motivations have been kind of turbocharged by the hype and the kind of promise of algorithmic applications all around.

And so in that environment, it becomes even more important to have basic guardrails around the purposes for which you can collect data, what are the legitimate grounds on which companies can collect data, how long can they keep it for, and what purposes can they use it for, right? If you don't have those guardrails, then I think we are going to see a sort of race to the bottom. And there's actually still no federal privacy framework even as the US is trying to kind of take leadership on AI.

RAFFI: Mm.

AMBA KAK: What I said in the testimony too that day, and I think a lot of us like you and others sort of echoed this as well was that, like data privacy law *is* AI law, like data privacy regulation is AI regulation.

RAFFI: We'll hear more from Amba later on this season, and we'll be talking a lot about regulation and legislation in future episodes.

Doing a podcast on data privacy alone would be *a lot*. But we're actually going to zoom out by asking what data even is.

For one thing, data, as a term, seems like it could refer to, like, anything. Or everything. Does all information count as "data"?

CHRIS WIGGINS: Generally about data, you're right that data can be like any ensemble of facts. It could be pictures. Could be numbers.

RAFFI: Chris Wiggins is an associate professor of applied mathematics at Columbia University. He also works as the chief data scientist at the New York Times.

CHRIS WIGGINS: I will tell you a linguistic fact that I like to hang on to, which is the etymology of the word data, which means something given. And I like that sort of vision that somebody just gave it to you.

Something becomes data the moment when you're like, I don't really wanna do a lot of critical inquiry into how these facts were created and what sort of subjective design choices went into it. I'm gonna reify it as data, as though it had some sort of objectivity to it.

RAFFI: Chris co-wrote a book that came out in 2023, along with his colleague, the historian of science Matthew L. Jones, who's now at Princeton. It's called *How Data Happened: A History from the Age of Reason to the Age of Algorithms.*

CHRIS WIGGINS: In our book, we make clear that data in the title, *How Data Happened*, really means data-empowered algorithms, deployed usually by private companies that are shaping our personal, political, and professional realities.

RAFFI: These data-powered algorithms are behind the scenes, using data about you — the people you follow, the purchases you've made, things you've Googled — all to shape the content you see. You know, for example, how Instagram decides which pictures to show you.

Or, a hiring manager could use an algorithm that takes what you've posted on social media and spits out a recommendation to hire you…ooooor *not* to hire you.

The point is, algorithms are fed by our personal data, the information that reveals important things about who we are and what we do.

And whether that data comes from us — entering info directly into a website, or if it's the result of some app tracking our clicks, likes, or purchases — we think our data is *out there somewhere*.

So Chris, sometimes you hear about our data being used to build quote unquote "models." It's not entirely clear what that means, but sometimes it seems like kind of a digital stand-in for us. Isn't that where things start to feel…icky?

CHRIS WIGGINS: Yeah, sure, that feels icky, but if it's just like, let's say my zip code, and there's a whole bunch of other people, particularly in my zip code, that have that zip code, then it somehow doesn't feel quite as icky.

I mean, for me, as somebody trained in mathematics, model is like an algebraic expression, right? Or in machine learning, a model would also be an algebraic expression, but represented on a computer somehow. But I can imagine "model" being something like a representation of you, and that representation could be very coarse or it could be very granular. And there's a point at which it becomes so granular that you start to feel icky about it.

RAFFI: Right, okay. I guess if you just have my ZIP code, that's one thing. But if someone has a whole list of things about me, like my ZIP code, but also my eye color, height, weight, web browsing history — what you just called a "granular" representation of me — that feels unnerving. That's a lot of information being captured without me realizing it. So like Chris, how do you think about this in your framework?

CHRIS WIGGINS: When somebody is capturing data about you and you're not really informed about it, then you haven't really given informed consent.

But I do think there are other things that are sort of icky about our feelings about privacy of data that are not exactly captured, which include, you know, somebody else is profiting off my data. And I feel like, how am I not profiting from my data, but somebody else is? That's sort of about fairness. Like somehow it doesn't seem right that these companies are making a lot of money off of my browsing data and all I got was you know free documents. It just doesn't seem fair that these companies are worth trillions of dollars and all I got was you know free music recommendations or something.

RAFFI: Chris is referring to the basic bargain of a lot of the internet right now. Whether it's Google Docs or Facebook: you can use the service for free…if you agree to hand over your data.

CLIP [CNBC; privacy policies]

> Anchor: *Data privacy has become a hot topic, which brings us to all of those privacy policies that we see when we log on to websites.*

RAFFI: Companies that collect our personal information have to make certain disclosures. And those are contained in corporate privacy policies. You know, those things that you click "Accept" on without reading them?

But I don't blame you. For one thing, you're in good company. According to a Pew survey from October 2023, 56 percent of Americans say they almost always or often click "Agree" *without* reading privacy policies. And between you and me? I bet it's even more of us.

And then there's the policies themselves. These are documents written by lawyers to protect the company. They're written to be cited in court, not to be *understood* by you.

And lastly, what choice do you have? If you don't click "Accept," you can't use whatever app or service you're trying to use. It's either.. say you agree, or nothing.

There's an old saying about how the internet works: "If the service is free, then you're the product." And that's one way to understand why companies like Google or Meta give away Gmail and Facebook for free. We sort of pay for them with data, rather than money.

But maybe there's another way to look at it.

> **CLIP:** [Shoshana Zuboff] *We thought these services were free. But actually, these companies think that we are free.*

RAFFI: That is Shoshana Zuboff, speaking here in 2019, at the Institute of Art and Ideas' "How The Light Gets In" festival.

> **CLIP: [Shoshana Zuboff]** *We think we're using social media, but actually, social media is using us.*

RAFFI: Zuboff is a professor emerita at the Harvard Business School, and the author of an extremely influential book on data privacy and the data economy called *The Age of Surveillance Capitalism*.

> **CLIP: [Shoshana Zuboff]** *Surveillance capitalism was a breakthrough idea where it was discovered that it was possible to secretly capture private human experience, and treat it as free, raw material for the translation into behavioral data.*

RAFFI: So, it's not even that you're the product. You're the raw material that companies use to make their products, which they then turn around and sell back to you.

> **CLIP: [Shoshana Zuboff]** *Our private experience has been commodified in the form of this behavioral data.*

ETHAN ZUCKERMAN: So we're all a little indebted to Shoshana Zuboff, who put a name on this. She calls it the ‘surveillance economy’.

RAFFI: This is Ethan Zuckerman, a professor at UMass Amherst, a longtime technology writer, and a key player in the history of the internet. We’ll hear a little more about that part later on.

ETHAN ZUCKERMAN: Her book, you know, *The Age of Surveillance Capitalism*, is this sort of thick tome, explaining the very complicated dynamics that happen when you know you're under surveillance all the time.

I like to teach my students about surveillance by getting them to fool around with TikTok. So you want TikTok to give you the content you want. You don't want it to give content you're not interested in. And TikTok is constantly collecting information on you. It's collecting how many seconds did you spend on something? Did you tap it so that you like stopped it and restarted it? Did you share it with someone? Did you like it?

Every one of those actions, even the passive ones, just sort of minutes of view time, end up being fed to TikTok so it can send you new stuff. To one extent or another, we're all doing that on all of the internet all the time. We have all inherited this world in which our actions online, particularly on our mobile phones, are being watched.

Some are very conscious actions. I subscribe to this person on Twitter or on Facebook. I bought a product, I signed up for a mailing list. A lot of them are much less conscious actions. I lingered a little longer on this story than on another story. All of that is being used as ad targeting data. All of that is being used as content targeting data, trying to gain more of your attention.

You don't really have a right to review this. Even if you could look at this data, it would be completely impenetrable. There's no way to understand how the algorithms are processing it. So we're living in this world where we have very limited control over how we're marketed to, over how content is targeted to us. And it is almost certainly changing our behavior. And frankly, most of us aren't even really aware.

RAFFI: So how do we talk about data flowing between, like, the people collecting it, the people who are reselling it, the people who are aggregating it, how should we talk about that?

ETHAN ZUCKERMAN: Yeah, so I think reselling is perhaps the key factor associated with it. The reason TikTok is able to give me content that's interesting to me is that they're sucking up a decent amount of my data.

And to a certain extent, I know and understand what that trade off is going to be. But then there's a second order effect. TikTok has a model of who that user is. And they know that user likes dogs. They know that user watches Mongolian cooking videos and enjoys watching people play tricky baselines on electric bass. What they do with that, I don't know. But they are almost certainly passing that data to partners and selling that data to data brokers. And that is going into some sort of a profile that people are using to target ads to me. And that feels a lot more complicated.

RAFFI: Data brokers are the unseen middlemen of the digital economy. They take personal information, from tons of sources, and turn them into products for sale to the many companies out there who are looking to buy them.

ETHAN ZUCKERMAN: So there are a number of apps that people have ended up putting on their phones that demand location data constantly, whether or not they need it. But their main purpose for it is to sell that location data to data brokers. I might be using app A, but I'm actually sending data to app B, which is selling it to broker C to sell it to advertiser D. Part of what's so tricky about this is this ecosystem is largely unregulated. It got built by the Googles and Facebooks of the world, but there are thousands of companies whose names we barely even know. And not all of them are doing things that we would probably be entirely ethically comfortable with.

RAFFI: The lack of transparency in these transactions can cause a lot of suspicion.

If you want the convenience of personalized apps, there's a chance that app might use your data in some unexpected way. Or sell it to some unknown company.

It's a tradeoff between privacy on the one hand, and basic functionality on the other. And it's the kind of thing we confront every day now.

ETHAN ZUCKERMAN: A lot of us are doing sort of calculations about what we do and don't share online, what we look for and don't look for, what different personae we're keeping to try to handle our relationships with the surveillance internet. And I suspect those subtle forces actually are extremely powerful and worth investigating.

RAFFI: This slider between privacy and convenience is just one of the many tradeoffs that we're gonna cover this season. The big question is: how are we supposed to live with all these tradeoffs?

> **CLIP: [Shoshana Zuboff]** *These behavioral data immediately, then, are declared as the private property of that corporation. . . .*

RAFFI: Again, Shoshana Zuboff.

> **CLIP: [Shoshana Zuboff]** *Now, with their private property, which is data about us, they can take that into their manufacturing processes — which are, of course, computational; we call it 'artificial intelligence.'*

RAFFI: So Chris, we're in this place right now where AI developers gather basically as much data as they can.

Chris: Yep.

RAFFI: And obviously that comes with a lot of issues. On the one hand, it's a security risk, right? If this data gets hacked or leaks, tons of stuff on tons of people is exposed.

But then also, why should they have *so much* data on us? Could we just limit the amount of information they're allowed to collect?

CHRIS WIGGINS: So there's two, I think, questions there. The first is how to square the language of mathematics with the language of ethics. The unit of analysis of ethics is decisions rather than models.

So somebody decided that the right use case for this piece of mathematics was to deploy it as a product that would then be used to inform hiring decisions. That's a decision we can analyze for its ethics. The decision to use this particular mathematical model in, for example, a judicial process, or hiring decisions, or any other way that algorithm and decision systems might be impacting people is the thing where I think we can actually perform a profitable ethical analysis. So, I mean, there's many, many ways that you can take something that is ethically challenged and try to make it, let's say, less evil.

So one way to make something less ethically challenged is to say, okay, I'm going to use as little data as possible. That has many benefits. But I think that the ways that you can make something less evil are much broader than using fewer data or using less granular information. That's nice, but there's all sorts of ways that you can prevent people from harms, that you can ensure justice, that you can ensure informed consent and respect for persons. There's many, many ways you can do that other than merely keeping fewer data.

RAFFI: I really like this way of thinking about data. It's the intersection of math, which treats data as rows on a spreadsheet, with ethics, because data comes from real people. And how we treat their data matters, because how we treat people matters.

So, Chris is right that limiting the amount of data is not gonna solve the ethical problem of eliminating harm. That's a much harder problem.

CHRIS WIGGINS: As a technologist and as a scientist, I think if you would have said to me five or six years ago, data has politics, I would have said, "Well, that's ridiculous."

Even in people deciding what data are to be kept and what data are to be thrown away has room for politics. And by politics, I don't mean voting. I mean of or relating to the dynamics of power.

RAFFI: The idea that data has politics is an old one. It might be at the forefront lately as we talk about AI, but it's a problem that's at least as old as the internet itself.

So we need to rewind and hear about the *early* internet — from some people who built it.

And that's what's coming up, after a short break.

[MIDROLL]

RAFFI: Welcome back to Technically Optimistic. I'm Raffi Krikorian.

LOU MONTULLI: The web started to grow very rapidly around the year 1993.

RAFFI: And this is Lou Montulli.

LOU MONTULLI: Certainly by the spring of '94, there were millions of people using the web. And it was all essentially just open source software written by college students.

RAFFI: And Lou here, he actually plays a really important role in this story. For one thing, he was on the founding team at Netscape. If you don't know, Netscape Navigator was one of the first mainstream web browsers. It was a big deal back in the nineties, and it had a huge part in shaping what the Web is today.

LOU MONTULLI: Netscape was formed in the spring of 1994. And we set out to solve the really monetary issue of the web. Because the web was built as open source software. There was really no thought given to how is anyone going to pay for this? Or what's the recurring revenue model of the web?

And so our concept was to build the software for free. The web browser would still be a free piece of software. But we would build the commercial underpinnings, i.e. the ability to do commerce securely through using encryption to build software that people could trust. So those applications would enable commerce, would enable revenue streams to be built. And Netscape's business model was to build the underlying infrastructure and server software that would enable all of this commercialization of the web and make money through that.

RAFFI: Here's a basic fact about how the web works. HTTP stands for Hypertext Transfer Protocol. You've seen it at the beginning of every web address, because It's the *protocol* of the web. It's how computers move data around the web. But HTTP is what they call a 'stateless protocol'.

LOU MONTULLI: The beautiful thing about a stateless protocol is that it's very efficient. You don't need to maintain a connection for each person that is using the web. You can merely connect, disconnect. Grab whatever data you need and then disconnect. And so you can literally have millions of people all sharing the same server. The downside is that a web server cannot tell the difference between a returning customer and somebody who is brand new at the time. So cookies were a solution to create an ability to remember a specific user in a stateless protocol.

RAFFI: That's right, this guy Lou Montulli, he invented the cookie. He holds the patent and everything.

LOU MONTULLI: And they were designed at the onset to be privacy protecting.

RAFFI: Even if you don't know what cookies do, you probably recognize the term from the annoying pop-ups you see all the time informing you that whatever website you're on uses cookies.

LOU MONTULLI: It's important to note that your browser doesn't necessarily know a lot about you as a person. And therefore, the website will not be able to know much about you as a person. But when your browser connects to the website for the first time, the website has the option to return a little ID that says, "Each time you come back here, give me back this same ID."

RAFFI: So, like, you're customer 10. Just keep on saying you're customer 10.

LOU MONTULLI: Exactly. So I connect to the web server. Web server says, "You're customer 10. Just tell me that each time you come back." And therefore the server can say, "Oh, customer 10, I know that previously you wanted to buy this one item. Would you still like to buy that?" So, for instance, it can keep it in a shopping cart for you to purchase later.

RAFFI: At its core, a cookie is a small piece of data. And as you browse the web, a site might give you a cookie: this little personalized chunk of data. Or it might ask to see the cookie it gave you last time. And it does this in order to give you a more personalized experience.

LOU MONTULLI: It took a while for the impact of cookies to really be felt. But within a year, there was cookies being used in virtually every site that did anything complicated.

RAFFI: Every cookie is tied to an individual web browser on an individual website. It's really Important to emphasize this point: cookies were originally designed to *protect* users' privacy.

LOU MONTULLI: Many people wanted to just add a unique identifier to the web browser.

RAFFI: So each web browser out there would just have a different ID, effectively.

LOU MONTULLI: Yeah. And the problem with that is that it would make tracking across many sites trivially easy. It's a bit like a license plate on your car.

RAFFI: Yeah, that's a good metaphor.

LOU MONTULLI: It's on there everywhere you go. You're known as that ID, so it's hard to get away from. We had a strong feeling within the web developer community that a user should be anonymous if they wanna be anonymous. And so we felt that we wanted to preserve privacy wherever we could.

RAFFI: So what went wrong with this implementation?

LOU MONTULLI: Yeah…So the most important one, and the one that most people, the reason most people know what cookies is, is that cookies became a core part of the ad tracking world.

RAFFI: The world of ad tracking.

> **CLIP: [Anchor]** *Trackers are companies you've never heard of actually, most of them. I mean they're small companies whose business it is to put little tiny things of software on websites, and those send back information about you and they compile a profile, and it's not your name usually, but…*

LOU MONTULLI: And the ad tracking world uses a specific type of cookie called a third-party cookie. And that is a cookie that is not associated directly to the website that you were intending to visit. You didn't explicitly visit this third website. You were going one place, but that website had a reference to this other website by an image, and the web browser went and automatically got that.

RAFFI: Third-party cookies enabled the modern ad tracking practices that dominate the web today. If you visit a website, it might download an image from some *other* web site. And this image, downloaded from a site you're not currently visiting, might pass along a third-party cookie.

And these images can be really small, literally one pixel by one pixel.

In fact, they're called tracking pixels. They're designed to hide in plain sight. And they're used by Google, Amazon, and Meta, and tons of other companies that use Google, Amazon, or Meta to advertise. When it's downloaded, it passes along a cookie that's used to track your browsing activity, and show you ads, all across the web.

This is why, if you spend time shopping online for a coffee maker, the ads you see on all sorts of different websites show you…coffeemakers. It's coffee makers everywhere.

The idea of some massive network of websites that all agree to track users in order to serve ads for *one* centralized platform — this was basically unfathomable back in the nineties.

LOU MONTULLI: The personalized advertising and ad tracking is actually an even more specific case. The ad tracking networks are really a, I don't want to use the word 'conspiracy', but it's a, it's, you have to, basically many, many sites have to conspire together to work with the same advertising network in order for ad tracking to actually happen. And so the way this works is, a particular ad tracking company and there really only been a handful of these that have ever really existed at scale. They contract with many, usually hundreds of websites to host advertising for them. It is only then possible that that advertising company could then see that the same person has gone to many different websites. So it's in that situation that you're giving away some, what I would consider private information, private web browsing information, to the secondary ad company. And that was certainly an unintended consequence of cookies as they were designed.

RAFFI: Yeah. There's like such an amazing contradiction here of, just like — you created a system that is attempting to be privacy preserving.

LOU MONTULLI: Mm-hmm.

RAFFI: And somehow it became an instrument to, like, compromise people's privacy in a lot of ways.

LOU MONTULLI: Exactly.

RAFFI: How did you realize that was happening?

LOU MONTULLI: Yeah, so this is somewhat of a pivotal moment in the cookie story. Around about 1996 it was made aware that this company DoubleClick was using cookies in an ad tracking network and they had the ability to track cookies across dozens, if not hundreds, of websites. They were still relatively small at the time, but they were a growing company, and they were generating a fair amount of ad revenue. And there was several interesting things about this. One, this shouldn't be possible. And it really was very annoying. Why is this happening? Cookies should not allow this.

RAFFI: Like, what's going on?

LOU MONTULLI: Because third-party cookies are interacting with image references, all these websites were conspiring together to work with the same company. It was a business model that we never really predicted. In this case, they were conspiring to try to advertise. And the data leakage, the privacy leakage, was an unintended consequence, a negative externality, if you will.

Advertising personalization was massively more effective than advertising without personalization. So by knowing a browsing history, you can quickly infer some very basic things about who might be behind the browser. So you might determine that this is likely a male visitor or a female visitor. You might likely know that they are from a wealthy area or a non wealthy area. And you can tailor your advertising towards those very rough cohorts. And so personalized advertising is somewhere on the effect of 10X more effective than non-personalized advertising.

RAFFI: Ok so just to, like, recap a second Lou. This thing you designed to protect people's privacy…it's now being used to sell ads. So like what was your reaction to this at the time? How did you respond?

LOU MONTULLI: Yeah, so the options on the table were, one of them would be to do nothing, would say, okay, this is a type of leak that is maybe not desirable, but also not disastrous. And we could just say, okay, that is okay. I didn't like that idea, but certain people did. Certainly the advertising companies like that idea.

RAFFI: Yeah.

LOU MONTULLI: The other idea in the opposite effect is to just disable third-party cookies. Now there are many legitimate uses for them. So for instance, comments on most websites are driven by a third party. It's completely legitimate. So disabling third-party cookies would have the effect of removing functionality that would otherwise be legitimate, as well as you would decimate online advertising for a large number of websites.

These cookie trackers were really solely responsible for generating revenue for a very large number of small to medium sized websites. And they still are today. Most websites who do not have hundreds of millions of users are using ad networks to become part of a larger whole so that they don't have to run their own advertising campaigns and have salespeople and all that.

And without these third-party advertising networks, these websites would have no revenue. If we turn off cookies entirely, we would be shutting off revenue to something like 70% of the internet, right? And so we worked really hard to try to find the middle ground because I didn't like either of the extremes. But we came up with a viable third option.

RAFFI: Mm-hmm.

LOU MONTULLI: And the option that we decided to go with was to provide the user with both visibility and controls to cookies that would allow them to notice when they were being used and to very explicitly set their own preferences. Both generally, like they could turn off all third party cookies, or they could do it, just on a very specific basis.

RAFFI: Earlier, I mentioned that you've probably seen cookie-related pop-ups on the web — the ones that say things like, "Hey, this website uses cookies, do you accept?" and then you consent to different categories of cookies. Like, strictly necessary, or marketing. These notifications mostly came about as a result of an EU law: the General Data Protection Regulation, or GDPR, which went into effect in 2018.

> **CLIP [CNN]:** *Well the privacy laws, as we've just been discussing, are about to come into force in the EU. They're gonna start a revolution in how companies handle personal data. Shelly Palmer is a technology expert…*

RAFFI: And even in the US, more people are talking about the idea of disabling third-party cookies by default. Google began rolling out this feature to some Chrome users back in January. That's actually something that Lou had been thinking about doing decades ago.

LOU MONTULLI: We gave the power back to the users, but didn't change the default. And that's maybe the most controversial part. So essentially, if we change the default to be don't accept third-party cookies, and still let the user turn them on, 99% of the people are not going to turn them on. And it essentially does the same effect as disabling all advertising.

RAFFI: Third-party cookies that track you all across the web are turned on by default. But most web browsers today already allow you to disable them. You can dive into the settings and do that right…now.

I'll give you a second.

Are you doing it?

Wait…you're not?

Ah. It's hard to convince people to make changes. Not everyone wants to tweak their browser settings. They're complicated, and people are afraid they might mess something up in the process. Plus, even if you know what you're doing…it's just kind of annoying.

So, if third-party cookies were turned *off* by default, it would be a big deal.

But there's a risk there, too. Web developers have built a ton of features that rely on third-party cookies. So if everyone were to suddenly disable them, it's not just ad tracking that would go away. We don't actually know all the things that might break around the web.

How are people supposed to navigate this by themselves?

JONATHAN ZITTRAIN: I think it is frankly too much to ask of regular people to make sense of it.

RAFFI: This is Jonathan Zittrain.

JONATHAN ZITTRAIN: And it's not to say you shouldn't try. And there can be guideposts online.

RAFFI: In addition to being a professor of both computer science and public policy at Harvard, Jonathan is on the board of the EFF.

JONATHAN ZITTRAIN: The Electronic Frontier Foundation, which is one of many organizations that offer, whether for journalists or others in sensitive professions or just the public at large, some advice on how to kind of batten down the hatches of your digital life.

RAFFI: Huh.

JONATHAN ZITTRAIN: And I think it's good practice to try to hew to that. But it's really tough. I think it's incumbent on us to figure out, systemically, what sort of privacy protections are important in a free, thriving society, and not to lean too hard on either user education or what is often referred to as user choice. Well, they opt into this, they opt out of that.

RAFFI: This user choice thing? It is a big issue. How much responsibility should be placed on you, the individual, to be aware of, or even accountable for, how your data is being used? Do you know what you've even agreed to in these privacy policies? Is the best shot you have at protecting your data you going deep into your phone's menus, turning off individual permissions one by one?

Because — let's be real — you're probably already letting apps do quite a bit behind the scenes.

JONATHAN ZITTRAIN: Regulating the use of cookies online to the extent that when you visit a website it now says, "Newsflash we use cookies! What would you like to do next?"

RAFFI: Jonathan's talking about those GDPR cookie pop-ups again. He's not a fan.

JONATHAN ZITTRAIN: And I mean the idea that, like, we need to educate the citizens of people about, like, what are these cookies? What are you talking about? Like only the strictly necessary cookies, please. Like, what does that even mean?

We shouldn't take much solace in the idea that people are making some kind of informed choice about it, if the dialogue box that you're presenting them amounts to, "Would you like to be taken advantage of, yes or no?" You should not be presenting that box, you should just not be taking advantage of them.

RAFFI: We didn’t opt in to any plan to extract and sell our data. And we certainly didn’t sign up to be taken advantage of. So, why should it be on us to opt out?

We’re gonna keep coming back to this question throughout the season. And you’ll hear more from Jonathan in future episodes.

LOU MONTULLI: So up until the last three or four years, cookies have been the primary means of ad tracking.

RAFFI: Here’s Lou Montulli again.

LOU MONTULLI: And what that means is, all of the advertising you see on the web is being served with a cookie. The great thing about cookies being the center of that is the user has control over whether or not they want to have those cookies cleared, if they want to opt out completely. They have one place to go where they can say, I want to opt out, I want to turn off third-party cookies, I never want to be part of this ever again.

RAFFI: For Lou, keeping users in control has always been a key part of cookie technology. And that is really a core value for him. But I guess at this point, we can see that this is a double-edged sword.

Back when third-party cookies first showed up, Lou had a choice to make. And today, decades later, developers are starting to look towards a possible future, beyond cookies.

LOU MONTULLI: Many other means of tracking exist. So one of the common ones is called fingerprinting. It takes a combination of your IP address and various other aspects of your web browser, and it combines that data to create a nearly unique token that basically represents the cookie again. The downside of it is you can never clear it because your IP address, your browser, those are all unchangeable characteristics. And so now you have a situation where you took away the ability of the user to opt out. You took away the user's ability to clear their history. So…we're moving from a world in which we had a set of tools that were really clear and also an ecosystem of other tools that allowed the user to become anonymous again.

I think in order to change these larger sociological problems, though, we may need to change the broader market in ways that individual engineers rarely have the capability to do so. Yeah, these are not necessarily technical problems. They're really sociological or legislative problems. If we legally support advertising and that is the revenue model of the web, then you're not going to be able to get rid of ad tracking. The technology will continue to evolve. And you may have browser companies continue to try to bring back more privacy, but those advertising companies will do everything they can to continue the ad tracking.

It's an important sociological question: Do we want this to happen or not?

RAFFI: Do we want this to happen…or not? This question is at the center of our show this season. Look, it's really easy, almost too easy, to feel very powerless in the face of technology. And that can make securing your personal data feel endless. And daunting. There's so little transparency into the data economy. So little light is shed on where your data goes once it leaves your phone and web browser and is out there in the world, powering ads, market research, and…who knows what else.

So, do we want this to happen, or not? We want smartphones that tell us about the traffic on the way to work, we want smart movie recommendations, we want personalized health care. We want all the many conveniences and quality of life improvements that data can give us. But can we have those things…and control our own data?

Here's the thing. At Technically Optimistic, our take is: Yes, we can. Just because tech companies have all your data now, and just because it all seems really complicated, doesn't mean that things can't drastically change. You have power. And I really look forward to showing you why I think so.

So next time, in episode 2, we're taking on social media.

We're exploring how just a few companies changed the whole landscape of online advertising…

> **CLIP: [Ethan Zuckerman]** *You know, those are the companies that really dominate the ad world. This has been a huge problem.*

How they got so powerful…

> **CLIP: [Meredith Whittaker]** *They collect huge amounts of metadata. They know who you're messaging, when. They have models of who you are and how you behave.*

… and how dangerous they *know* they've become, as we're joined by Frances Haugen, the Facebook whistleblower.

> **CLIP: [Frances Haugen].** *Facebook knew these problems were real. They told the public they were not real.*

That's next time… on Technically Optimistic.

CREDITS:

Technically Optimistic is produced by Emerson Collective. Production assistance from Christine Muhlke. With original music by Mattie Safer. Our senior producer is Erikk Geannikis.

If you've enjoyed the podcast, please rate and review us – and subscribe on Apple, Spotify, or wherever you get your podcasts. Follow along on social, @emersoncollective.

And sign up for the Technically Optimistic newsletter! You'll get my thoughts about the week in tech, with lots of big questions, interesting links, and, you know, tons of ways to get in on the conversation. Subscribe for free at technically optimistic.substack.com.

I'm Raffi Krikorian. Thanks for listening. See you next time.