[DAISY THE GREAT](https://www.facebook.com/daisythegreatband/) [BUILT MY HOME ON HOLLOW GROUND](https://www.youtube.com/watch?v=z-1sC1lkmKw)

CRAIG: Hi this is [Craig Smith](https://en.wikipedia.org/wiki/Craig_S._Smith) with [Eye on AI](https://www.eye-on.ai/podcast-archive), a podcast about artificial intelligence. [Data is valuable](https://www.nytimes.com/interactive/2019/09/23/opinion/data-privacy-jaron-lanier.html) – something that companies like Facebook, Google and Amazon realized far earlier than most consumers did. Most of us have given away our data, signing it over to companies with the click of an online ‘Accept’ button, not even bothering to read the fine print. Why bother? Either we accept the terms and participate in the digital world or we unplug – something that’s not really an option for anyone operating in the global economy. Fortunes were built on that data, enriching [a handful of entrepreneurs](https://www.entrepreneur.com/article/319952) to heretofore unheard-of heights.

But computer scientists have been working on alternative models, even as the public has grown weary of having their data used and abused. This week, I talk to [Dawn Song](https://en.wikipedia.org/wiki/Dawn_Song), one of the world’s foremost experts in computer security, about her vision of a new paradigm in which people control their data and are compensated for its use by corporations. Dawn, a professor at the University of California, Berkeley, has recently launched a company, [Oasis Labs](https://www.oasislabs.com/), which is building a platform that brings together the immutability and integrity of blockchain and the security and privacy offered [by trusted execution environments](https://www.trustonic.com/news/technology/what-is-a-trusted-execution-environment-tee/), among other technologies, to give data owners the ability to control their data and audit how it is used. On top of that, Dawn and her colleagues have proposed a decentralized, [federated learning](https://ai.googleblog.com/2017/04/federated-learning-collaborative.html) model called [HiveMind](https://devcon.org/agenda?talk=recxSp8jJIzeusvEW), which allows engineers to train machine-learning models on the platform without ever seeing the training data.

I hope you find Dawn’s ideas as revolutionary as I did.

CRAIG: Could you start by introducing yourself and giving us a sense of where you're from. I know that you went to Tsinghua for undergraduate.

DAWN: For undergraduate. Yes, yes. Yeah. My name is Dawn Song. I'm a professor in computer science at UC Berkeley. I've been a professor here for more than 12 years now. So, I got my undergraduate degree from Tsinghua actually in physics, and then when I first came to the United States, I was in the PhD program at Cornell for physics. After one year I decided to switch to computer science, so I went to Carnegie Mellon in computer science and in the end, I actually got my PhD from UC Berkeley.

CRAIG: Wow.

DAWN: Yeah. It was a really great experience. I got to know, you know, more faculty and have more great fellow students during the PhD time.

CRAIG: Where did you grow up in China and I'm curious what your parents did. I'm always curious about what kinds of households people grew up in that got them into science.

DAWN: I grew up in [Dalian](https://en.wikipedia.org/wiki/Dalian). It’s in the northeast part of China.

CRAIG: I know Dalian well.

DAWN: Yeah, it's a beautiful city. It's actually very similar to San Francisco in many aspects. It's at the tip of a peninsula and also it has hills and has these beautiful cliffs, you can drive along, like a highway, along the cliffs. Yeah, very pretty. My parents didn’t really do science or anything. They were in the government, just normal employees.

CRAIG: You didn't grow up in an academic household?

No. No.

So, you went to Tsinghua knowing that you were going to study physics?

DAWN: Yes, actually it was a, it was a funny story. I always loved science. I always did well in math and physics and when I was applying for college - so everybody goes through the same [national entrance exam](http://www.thatsmags.com/china/post/13965/explainer-gaokao). And then before the exam you have to put in your application, which schools and what you want to study. And actually, at the time my mom wanted me to study business and she had put in an application for me for one of the business schools. And then it was my high school teacher - I really appreciate him, he has been actually really influential in my life. So, he was the one who actually told my mom that I have great talents for science. He thinks that if I study science I can't have a better future that is more suited for me. And then he was able to convince my mom. So, in the end my mom changed my application and then I applied to Tsinghua.

CRAIG: Oh, that's wonderful.

DAWN: Yeah, actually when I was in high school, at one point I actually wanted to become a photographer.

CRAIG: Oh really?

DAWN: More like a National Geographic type of photographer, going into the wilderness and also it was my high school teacher who convinced me that that's not suited for me. My talents were in science.

CRAIG: So, if it weren't for him, you might be a struggling photographer earning very little money,

DAWN: Possibly, yeah.

CRAIG: Then the PhD at Berkeley, what was your thesis?

DAWN: My main PhD work has been in computer security. So, in particular the thesis was actually developing techniques and ultimately the tools for analyzing security protocols and systems. And the nice thing about security is that it's so interdisciplinary and it's really about the security of different types of systems. So earlier we talked about security of the network, security of open systems, security of the web and mobile applications. And then when you talk about machine learning, machine learning systems and so on, then of course you need to think about how attackers may attack these machine learning AI systems as well. So, it's very natural to think about the adversarial aspects.

CRAIG: The PhD was not specifically in, in machine learning or AI, it was broader than that.

DAWN: It was more in security. Although even during my PhD time, I've found some work actually at the intersection between security and AI, machine learning. Even people today, sometimes when they meet me, they say that they remember that work. Using machine learning you can, essentially from [keystroke timing](https://arxiv.org/pdf/1706.06381.pdf), you can infer what people are typing.

Wow.

Using [HMM models](https://en.wikipedia.org/wiki/Hidden_Markov_model) and then you can actually get this keystroke timing from eavesdropping on the network. And from there you essentially can infer the keys that people are typing.

CRAIG: And, and so there's a kind of interplay or a dialogue between people that study adversarial attacks and people that come up with security solutions.

DAWN: Yeah. Because the attackers are always trying to evade detection, evade these defense systems.

CRAIG: At this point, how secure do you feel the standard algorithmic models are, I mean, convolutional networks or NLP solutions, from adversarial attack? I mean, have they developed or they still very vulnerable.

DAWN: Today, unfortunately, we still don't have very robust, resilient, learning systems that can be really effective against these attacks. So, my work and also other researchers work has been working on developing better learning systems that can be more robust. But today still, you know, strong adaptive attacks can still be very effective in most cases. Researching computer security, I not only study attacks, but really that’s to help us to understand more the limitations of the systems that we want to protect. And the ultimate goal is always to build more secure and more robust systems that can defend against attacks. And so, for learning systems, as we deploy these learning systems more broadly, then it's even more important that we make sure that these learning systems are secure and robust against these attacks. So, it's, it's very natural and it's very important to build better defenses.

MUSIC: INTERLUDE

CRAIG: A few years ago, you published a paper on, on a platform called [Ekiden](https://arxiv.org/abs/1804.05141).

DAWN: That's on blockchain.

CRAIG: That's right. That marries blockchain with secure enclaves. Is that the basic idea behind Ekiden?

DAWN: Yes. Right.

CRAIG: Did that grow out of this work on trying to find more secure solutions for computer systems generally?

DAWN: Yes. You know, my work in computer security, I've been trying to develop techniques and systems that can provide both security and privacy solutions. So, in this case, by marrying secure enclaves and blockchain, essentially the hope is that you get the better of both worlds. So that you can actually have the integrity that the blockchain ledger provides and also you can have privacy or confidentiality for the smart contract execution that's provided by the secure enclave.

CRAIG: Yeah. Because what I'm interested in is the evolution of computer security and then you develop the Ekiden platform idea. And then from there you launched Oasis. So how does all that fit together?

DAWN: I mean, in general, for my research, I work on different projects, Ekiden is one of the projects and HiveMind is another project, and they solve different problems. In this case, HiveMind also utilizes some of this decentralization blockchain technology that we have built. But more specifically for doing this decentralized training that doesn't rely on any central trust.

CRAIG: So they are separate research projects, but related. Are those the two core ideas behind the Oasis platform?

DAWN: So the Oasis platform, essentially, we built a blockchain smart contract platform that provides greater scalability and privacy protection. And on top of that we enable these privacy-preserving applications to help users maintain control of their data and their rights to data. And at the same time to enable data to be utilized in a privacy-preserving way. For example, for privacy preserving analytics and machine learning and so on. So, in this case, essentially, that privacy preserving machine learning is an application built and enabled on top. Privacy preserving machine learning is one application domain that we enable using the technology that we have.

MUSIC: INTERLUDE

CRAIG: Last time we talked you were talking about how data privacy or data ownership is a new idea in society and how, once people recognize that data is property …

DAWN: Yeah. This can you enable new business models and also essentially can, I think, enable new growth in the economy. Just like in the past, early on, we didn’t have a notion of property rights and we didn't have this notion like property belongs to XYZ. But once [property rights were institutionalized](https://en.wikipedia.org/wiki/Right_to_property), then it really launched the economy into a new boom. And similarly, we hope that, going forward, this data rights, that data can be viewed as property, can also essentially help propel the economy into a new level. And that new business models can be built on this and users will benefit from that. But, of course, data is different from physical objects. For example, data has this property called non-rivalry. It’s not like a physical object. If I give you an apple, then someone else cannot have the apple, only you can have this apple. But with data, if I gave you a copy, I can still give another copy of the data to somebody else. With data rights, hopefully, even though there are these non-rivalry properties, but with data rights still we can help users maintain control of their rights to data and users can still gain benefits from that.

CRAIG: So part of your research has been in looking for solutions to both protect data and be able to use data in secure ways.

DAWN: Right. I think going forward, that's really, I think that's really the key, right? It's really important. The whole world is collecting more and more data and a lot of the data is sensitive. So, either you lock data up, the data stays in silos and then the data won't be utilized, or if you try to utilize the data, then you need to make sure that the, you know, data privacy is being protected and so on. So, I think going forwards, especially as we talk about data is the new oil we are getting into the information age and all that, it's particularly important to develop technologies that can utilize data in a privacy-preserving way, in a responsible and compliant way. Blockchain is, because of this immutability and integrity guarantees and all that, so blockchain is a great way to essentially record and audit, for example, your rights, including your rights to data.

CRAIG: The model so far, and the reason that society is having such a problem with this, is data is collected and stored on centralized servers that are vulnerable to attack. And the advantage of blockchain is that there's no centralized server, that it's dispersed. The problem with blockchain is that the data is by nature public, by nature of the blockchain. So, what you're doing is taking the blockchain model but then creating a privacy layer.

DAWN: Yeah. A confidential computing layer.

CRAIG: Yeah. On that blockchain. So, Ekiden, was that kind of the first step in that?

DAWN: Yeah, you can view it that way. You can view that, I think that was the first paper that could show by combining secure enclave and blockchain, essentially you can get properties from both.

CRAIG: By adding federated learning, which you proposed in this paper called HiveMind ...

DAWN: So HiveMind is a different paradigm. So what HiveMind is trying to enable is this decentralized machine learning and machine training. We have seen federated learning where so far like Google and Apple have deployed federated learning so that the raw data, for example, stays on individual user's device. So, the central server doesn't actually ever see the raw data. And still you want to use all this data to train a machine learning model. So how it works is that the central server in this case coordinates all these different individual users' devices and have each user's device send what's called gradient updates from using the local data to train the model to the server and then the server can aggregate these gradient updates, and this is called federated learning. Although in federated learning you still have a central server who is doing the coordination. So, you're still relying on the central server in this case. But in the decentralized application case, we don't want to have to rely on any central server. So then in this case, for example, on the Oasis platform, we can actually enable this type of smart contract in a decentralized manner. We can have this smart contract coordinate the distributed or decentralized machine learning instead of relying on a central server. So, we can enable these different devices to together train a machine learning model. But you don't need to rely on any central server.

CRAIG: And Oasis, the Oasis platform is bringing together a lot of these different technologies. One of them is the secure enclave idea.

DAWN: Right. Which enables confidential smart contracts.

CRAIG: That's right. To run on the blockchain. There is then also this idea of being able to use the platform for

DAWN: decentralized training.

CRAIG: Yeah. Federated Learning.

DAWN: Right, right.

CRAIG: And, and where does the platform stand now? I mean, have you, do you have all of these pieces put together yet?

DAWN: Oh, we just released our [DevNet 2.0](https://medium.com/oasislabs/devnet-2-0-and-our-new-oasis-sdk-c858c25716e7) so developers today can, can go to our website and try it out and on the DevNet they can build now on our DevNet to build different applications.

MUSIC: INTERLUDE

CRAIG: If the platform develops critical mass so that you have enough nodes that it's working efficiently, can you relate that to some of the data breaches that we've seen recently and how those data breaches could have been prevented or avoided if there had been this kind of a platform available? I mean, for example, Equifax, the big [Equifax breach](https://www.equifaxbreachsettlement.com/).

DAWN: So, of course, computer security is a very complex problem. It's not just about whether it's centralized or decentralized. Decentralized just enables you to build systems that don't rely on the trust of any central party. With the traditional blockchain, all the data is public, so you can't really protect the privacy and confidentiality of that data. Data breaches can happen for many different reasons. So, in Equifax' case, actually there have been security vulnerabilities in the application. So, when you want to protect the security and privacy of data and still be able to utilize it in a secure and privacy preserving way, then there are many things you need to do. You need to protect the, essentially the computation process from leaking sensitivity information. So that's where the secure enclave can be helpful. Essentially a secure enclave provides a practical way for doing confidential computing where the data outside of the secure enclave is encrypted.

DAWN: It's only decrypted inside the secret enclave and a program gets executed inside the secure enclave to process the data. So, this way the computation process protects the sensitive data from being leaked or being stolen. But at the same time also you need to ensure that whatever the application does, it doesn't leak sensitive information. So, for example, you can have a program running inside the secure enclave, but you also need to ensure that the application itself doesn't leak sensitivity information. And then depending on what the application does, then you can use different techniques to ensure that. For example, one of my recent works in collaboration with researchers from Google and so on looked at even when your train a machine-learning model, the [machine learning model itself could be remembering the sensitive information from the training data](https://arxiv.org/pdf/1802.08232.pdf) and then an attacker, from just querying the machine learning model, could actually extract sensitivity information that were included in the original training data.

DAWN: For example, credit card numbers and social security numbers that were included in the original training data while you're training a language model. And our process showed that instead of training a vanilla machine learning model, if you're training a differential private language model in this case, then this differential private language model actually can provide much higher privacy protection for the original training data. So, this is an example showing that even as you run an application on sensitive data, you need to protect the outputs of the application from the leaking sensitive information. So, for example, as you train machine learning models, it's beneficial to provide these differentially privates solution. And similarly, for data analytics as well. So, for example, some of our technology has been deployed at Uber to enable privacy preserving data analytics incorporating differential privacy. So, this way it protects the query results from leaking sensitive information about the original data. If you actually use methodologies and mechanisms that can analyze these applications, and enforce the applications from leaking sensitive information, then certain attacks can be prevented.

CRAIG: Yeah. And how then does, does the Oasis platform, how do enterprises in particular use it or application developers use it so that their applications are secure from those kinds of breaches?

DAWN: Right. So essentially, we combine a number of different technologies. In our platform, the data is stored in encrypted form. With the data, we also include policies that specify how the data can be utilized and also include information about data ownership and so on. And these policies and the records of data ownership is stored in the ledger so that it's immutable and has high integrity, and all that. And then, if an application, or program wants to access data, first the policy in the smart contract, policy agents, will first check to make sure that the application actually can have access to the data, can utilize the data. So, the use of data will be enforced to ensure that it is compliant with the policy of the data, and then the application itself will be run in a secure execution environment to protect the computation process from leaking sensitive information. So essentially, we provide end to end security protection for the entire data life cycle.

CRAIG: And, the applications would be built on this platform.

DAWN: The application will run on the platform.

CRAIG: Run on the platform,

DAWN: But this enforcement, to check whether the application complies with the policy, this part applies to certain types of applications. So, for example, if you want train a machine learning model we can ensure, that the model is differentially private. So that you can make sure that the resulting model doesn't leak sensitive information.

MUSIC: INTERLUDE

CRAIG: Can you relate this, the platform and applications that run on the platform, to individual data rights? I mean, how does that translate into an individual understanding that ...

DAWN: That's a very good question. So, for example, if your train a machine-learning model, the machine-learning model may use different data points in the training data set and each data point may come from different users. And in this case, we can actually know and record that this machine-learning model, actually is trained from these different users' data. So, one work that we did recently, was to actually show how we can evaluate the contribution of each data point to the final training of the machine-learning model. So, in a sense to do, we call it data valuation and to essentially evaluate how important, how valuable a data point is to the final trained machine-learning model. We use a notion called Shapley value that essentially allows you to evaluate each data point's contribution to the model. And this way, also by using smart contracts, we can, essentially audit and log how the data is being utilized and how it contributes to the final model. And also, we can actually evaluate the importance and contribution of the data point.

CRAIG: You have a program right now with Stanford.

DAWN: Hospital.

CRAIG: the [Kara project](https://kara.cloud/#/). Can you describe that a little bit? Because that's what you're talking about, right, about people ...

DAWN: So that's one, that's one application. Right. We have a clinical study approved at Stanford Hospital in collaboration with the Stanford doctors utilizing our technology, essentially to help patients contribute their medical data for medical research. So essentially with our technology we help protect the privacy of patients' data and also at the same time to enable the data to be utilized in a privacy preserving way. For example, to train a privacy preserving machine-learning model and do analytics and so on to help medical researchers to find better cures for diseases.

CRAIG: And aren't you looking, or isn't Stanford looking at how they can incentivize patients to contribute their data?

DAWN: This is part of the study, too, is that we wanted to study to see what kind of incentivization the patients will find meaningful to them to contribute data for medical research and so on.

CRAIG: Yeah, and do you envision a day when applications will be built on a platform like Oasis so that individuals can manage how their data is used and also have an audit of how their data is used and be compensated for how their data is used? Do you envision that someday people will have an individual revenue stream from their data?

DAWN: Yeah, very likely. I really hope that in the future we can achieve something like this because an individual's data is valuable. This data, you know, essentially conveys a lot of information about each individual. A lot of services can really benefit from this data to provide better service to the user. With our technology, we really hope to enable this new paradigm where users can maintain control of their data and also at the same time they can enable certain applications to utilize the data, but the application can only use the data. The application cannot take a copy of the data out to do whatever the application wants. The application can only use the data, for example, to train a machine-learning model or to do something and also the user can provide terms. For example, if the application wants to use it, the application has to provide certain incentives to the user, something like that.

DAWN: That way essentially we are turning the world around. Today, obviously private services they are taking users data essentially using it as a product. They take users data and then they monetize it and users still get some service out of it. But, I think the world can be very different if this is turned around, where users get to maintain control of the data. So, in this world users don't have control of the data. The data is in third party services hands and they get to decide how they want to use it. But if users can maintain control of their data and applications can only use it in certain ways that the users agree and also users can gain potential revenue from their data. I think that can really enable a very different world.

CRAIG: Yeah. I guess the question on the benefit to the user is the companies that are using the data, it's valuable to them because they're collecting hundreds of millions of data points and there's value in that kind of an aggregation. An individual data point is not very valuable. If you take the revenue that that Facebook generates off of its use of the data that it collects, it's significant to a company like Facebook. But if you were to divide that among however many users Facebook has now, it would be an insignificant amount to each individual. Right?

DAWN: Yeah. So that's a, that's a very interesting question. How, how does revenue really get distributed? For example, if you want to train a machine-learning model, you maybe don't need like a billion users' data. Whoever contributes to the model, who has his data points contribute more to the training of the machine learning model can get more of the benefits. So, this is a different distribution model. And also, if you change the world into this different paradigm, then it's not just contributing to Facebook. Also, it's contributed to many other competitors of Facebook and many other services. And if each one gives you a small revenue stream, then now you are aggregating these small revenue streams and that could become something meaningful.

CRAIG: Yeah, it could become something over the course of a lifetime. It may not be something significant on a monthly basis, or annual basis, but maybe it could contribute to your retirement. The fees that you've accumulated over the course of a lifetime for companies using your data aggregated could be something significant.

DAWN: And also this way potentially you may even get better services. Because now, they get access to our data to provide services to you but in a more controlled and privacy presuming way. So, in the end users could get better services as well.

MUSIC: INTERLUDE

CRAIG: So the platform that you've built, or you're building with Oasis, what's the primary problem that you're solving? Is it to give individuals better control and potential benefits from their data? Or is it to provide developers and enterprises a solution to protect the data the that they're aggregating?

DAWN: So ultimately I think we want to provide the technology and to help like developers and enterprises to build these secure and privacy-preserving applications and systems to enable users to then protect privacy of users data and still at the same time to enable data to be utilized the privacy, preserving way

CRAIG: the trusted execution enclave, TEE, and the federated learning, these are things that have been in the air for a number of years now. Are there other systems that have put all of these pieces together? What is it that makes the Oasis platform unique?

DAWN: I'd say really that we combine these technologies to really enable this end to end secure protection of the data life cycle.

CRAIG: Yeah. What's the next milestone?

DAWN: We released our DevNet 2.0. And then we hope to have more applications built on top of the platform.

CRAIG: And the applications that are being built on the Dev net will be commercially available applications or are they research projects?

DAWN: I mean, right now we are building products, but the hope is that these will become commercially available. We are exploring the healthcare space and also the financial space for financial services and so on. I don't think we're ready to talk about it yet. So, talking about secure enclave, it's really important to build a truly secure enclave. And so today, the commercial solutions that are available, they are closed sources, so it's very difficult to know how secure they are. So, at Berkeley, we have been doing a research project called Keystone and developing an open source secure enclave. It's also built on top of Risk5, which is open source risk architecture. So, this way we can have the community come together to analyze and verify the correctness and the security of this open source secure enclave. So, we hope that this can really help us to build a more secure and truly trustworthy secure enclave.

CRAIG: That project would be open source, but then would be taken up by chip manufacturers? Because the secure enclave is a hardware solution.

DAWN: Yes. Right. So, yeah. So, we hope to also work with manufacturers and giving that to open stories at some point. Yeah. Anyone can keep manufacturers as well. That would be an important component in having the world shift to this blockchain, into this secure computing paradigm.

CRAIG: Are there any particular vulnerabilities to TEEs now?

DAWN: There have been, essentially, side channel type of attacks. But also, in general, we want to enable these secure enclaves that do not rely on the trust of a central entity.

CRAIG: That’s it for this week’s podcast. I want to thank Dawn for her time. For those of you who want to learn more about the things we talked about today, you can find a transcript of this episode on our website: Eye-on.ai. Let us know if you find the podcast interesting or useful and whether you have any suggestions about how we can improve. We need support to broaden our reach, so we’d really appreciate your rating or reviewing or sharing the episode.

The singularity may not be near, but AI is about to change your world, so pay attention.