



Podcast Transcript

Reimagining Data in the Wave of AI

Date: April 16, 2024

Guest: Janine Anthony Bowen, **Host:** Amy Kattman

Run Time: 17:12

For questions and comments contact:



Janine Anthony Bowen

Partner

Atlanta

T: +1.404.946.9816 | jbowen@bakerlaw.com

Kattman: Generative Artificial Intelligence tools, often abbreviated to Gen AI, require vast amounts of data to train their models. For companies offering Gen AI tools, it is common for those companies to use the data of their customers to feed and fine tune their models. Often the data rights are negotiated in the technology contract between the Gen AI provider and its customers. This approach to getting access to data, while not new, has significantly different ramifications in the Gen AI context and will require companies to rethink their approaches to data ownership, data licensing, and data classification. I'm Amy Kattman, and you're listening to BakerHosts.

On today's episode, we explore data-related issues and technology contracts in the age of Generative AI with Janine Anthony Bowen. Janine co-leads both the Privacy Governance and Technology Transactions and the Digital Transformation and Data Economy teams at BakerHostetler. Additionally, she serves as the Atlanta and Orlando Digital Assets and Data Management leader. Welcome to the show, Janine.

Bowen: Thank you. I'm glad to be here.

Kattman: Janine, Artificial Intelligence has brought the role and necessity of data into renewed focus. Can you tell us why it is important for companies to reconsider their approach to data in their technology vendor agreements?

Bowen: It is a perfect question and thank you for that. Well, interestingly, over time, vendors have been very deliberate about attempting to get access to data in the

technology contract. This has become even more acute in the age of Artificial Intelligence, especially with respect to Generative AI, in part because the large language models that underpin the Artificial Intelligence systems depend on vast quantities of data in order to deliver quality results. So, what that means is that vendors want to scoop as much of their customer's data into their processing and into their systems and into their Artificial Intelligence models than they have previously wanted to do. And so, customers now must pay attention to the way that they think about data and technology contracts.

There are a couple of different buckets that I'll just call out now and then we'll go into a little more deeply in a couple of minutes. So, in vendor agreements right now, all data is considered customer data. But there are multiple buckets. There is truly confidential information, there is business data, and then there is personal information. And each of those has different value to the business. Also, each of those has potential value to the vendor. And so, what Artificial Intelligence systems are looking for is mass quantities of information, and those vendor contracts represent the best opportunity for the vendors to gather it. So, that is why it is important for us to think about how we understand data in a vendor contract, because the customer is potentially relinquishing value to the vendor and the customer should be aware of the various different buckets where the value sits from a data perspective.

Kattman: Thanks, Janine. Could you provide us with a little history on how we come to understand data in tech contracts?

Bowen: Yeah. Alright, so let's go back a couple of generations. A couple of generations ago, the value in a business was in the real property. The building, the hardware, the products. Over time, the value in a business has moved from real property to intellectual property. That is trademarks, that is patents, that is copyrights, that is software, that is technology, that is intellectual assets that are not necessarily physical in nature. And so, two generations ago, real property was the value of the business. In the last generation, intellectual property was the value of the business. In this generation, data is the value of the business. And understanding that there is so much opportunity in understanding ways to release value associated with collected information.

So, we used to think of confidential information as business confidential information. Financial plans, trade secrets, marketing data, other items of competitive advantage that the business kept secret. And if that information was divulged, there was a clear negative consequence to the business. Well, as personal information started taking steam in the era of cloud computing, and regulation started to be introduced around the protection of personal data, it became convenient for the customer to say, not only is my business confidential information secret, but I also need you to protect my personal data as if it were confidential. Okay? And then there is this third bucket of data, and this third bucket is just what we'll call customer data. It is the data that the customer must provide the vendor in order for the vendor to do the processing and deliver the services that the customer is paying for. And so, what has happened over time is all three of those buckets, true business confidential information, personal

information, and all the other data that is being provided to deliver the service has all been considered confidential. But as a practical matter, they're not all confidential. Okay? And what we are now learning more deeply in the era of AI is, if a vendor wants to consume customer data for its own business purposes, it is going to be important to segment out those data types in the mind of the customer.

So, I'm concerned about confidential information for sure. Business confidential information that has this whole body of law around it and as the customer, I probably don't want that in the hands of a vendor. Okay? Personal data, which is also in this whole customer information or customer data bucket, has an entire regulatory scheme around it. But if you strip off the personal elements, there is value in that data as well. Actually, whether or not you strip off the personal elements. But for sure, the customer would have an objection to the vendor taking the personal data elements and creating additional products and services around that. And then you have just the run-of-the-mill data that is in there that also might be valuable to either feed the customer's AI systems, or the vendor's AI systems.

So, we can no longer treat them equally because the customer may be prepared to share general customer data that has been de-identified, anonymized, where you cannot tell where the source information came from, the customer might be prepared to allow the vendor access to that for the vendor's own business purposes, which would be different than the purpose of delivering the service. But the customer must now consider this more carefully and now break up these three data types again in the customer mindset to understand which piece of the value is going to allow the vendor to have access to, if any. Because it is not inappropriate for a customer to say you can't use any of my data. However, a vendor will say, we're all in this this community together and we all want to feed these AI models because there is going to be value in the technology that I give to you, or that I'm later able to commercialize later, if I have access to your data streams along with all my other customer data streams.

And so, understanding that we've moved from true confidential information that everyone agrees should be secret to all this other stuff that has been treated secret but is not necessarily, along with the vendor push to get access to it for purposes unrelated to the purpose that it was collected, gives us a very different way of thinking about the way that customer data should be treated in the hands of a vendor. So that is a mouthful, so I'll stop there, and we can tease that out more if we like. But we can probably move forward, too.

Kattman: Well, that was a really good description of the different data types. But could you talk now about what are the issues with our current approach to data?

Bowen: Okay. So, I've attempted to table set that customer thinking of all the data as confidential information is problematic in the age of AI, for reasons that we just discussed. There is value in each of those different buckets, true confidential, ordinary business information, personal data. But if they're all being fed into large language models, then things can happen.

So, for example, if true confidential information ends up as part of a large language model, then there has been the inadvertent loss of the confidential treatment of that data once it is consumed. So, if you're really concerned about protecting confidentiality, then the customer must be deliberate about what is confidential information that we are providing a vendor, and how do we prevent that data from being consumed in the AI system, in the Generative AI system? So, the inadvertent loss of confidential information treatment is a serious risk that we have to consider. Okay?

Now, the obligations associated with personal data still exist. But if a vendor wants to consume the personal information, then there are legacy obligations from a regulatory and compliance perspective that will attach that will make that potentially unpalatable to both the customer and to the vendor. But those get confused over time when you're trying to figure out how to segment out personal data or data that cannot be converted to information that is not personal for purposes of consumption by large language model.

The other issue that we need to think about is, there really could be very significant value, business value, in data that is neither proprietary, as in confidential, or personal. So, the rest of the stuff that is in there that we've been treating confidential over the last decade or so, that really is just business data, there is a lot of value to be unlocked there. And so, the customer has to get a handle on, how do I want to maximize the value of that data? Do I keep that internal? Do I share it out? And what is the business consequence or business value to me as the customer if I'm prepared to let vendors have access to that body of data outside of the reasons for which the contract was entered into?

So, the issues are, kind of recap, you can lose your confidential treatment, which is a problem. You can't let go of the obligations around personal data, which can get confused as vendors attempt to strip off the personal elements, and whether they do a good job of that or not. And then third, understanding that there is a lot of value in that third set, the non-confidential information, non-personal information piece that is remaining, and figuring out how to extract the value in a way that makes sense for the business.

Kattman: Well, you certainly have explained the concerns, but could you tell us how a company should proceed in light of all these concerns?

Bowen: Yes. So, three different ways that you have the customers or companies need to think about this from a commercial contracting perspective. First, I hope I've made the case that it is important to understand the value of each data type. And so that is the first thing, the first concern, or the first thing that a company should get a handle on. How do we value, or how do we see the value in the marketplace of these various types of data? And to be deliberate about understanding that.

Secondly, once the customer has gone through that process of understanding the value in each of these buckets, then from a vendor contracting perspective, the company needs to rethink how it considers the data in those agreements. So,

will it maintain this approach that it is all confidential and therefore subject to all the law around the release, inadvertent release of confidential information? Or will it take a different approach with respect to the three buckets, with respect to what it will provide to a vendor later on in the contracting process for the vendor's own subsequent use.

And finally, it is important that the company has a clear, rational approach to what data that it will share and won't share with third party vendors. The vendor will want it all. And the vendor will say, I'm not trying to take ownership of it, I'll just take a license. Or the vendor will say, I'm going to anonymize it, de-identify it, strip off any indication that it comes from you, customer, and then I want to own it. And so, company has to decide, is it comfortable with those various types of approaches? And how does it change its model around ownership and commercialization of its own data given what we see from vendors with respect to building out their own Artificial Intelligence systems?

Kattman: Janine, thank you. This has been very informative, and very helpful.

Bowen: Thank you. It has been my pleasure to be with you.

Kattman: If you have any questions for Janine, her contact information is in the show notes. As always, thanks for listening to BakerHosts.

Comments heard on BakerHosts are for informational purposes and should not be construed as legal advice regarding any specific facts or circumstances. Listeners should not act upon the information provided on BakerHosts without first consulting with a lawyer directly. The opinions expressed on BakerHosts are those of participants appearing on the program and do not necessarily reflect those of the firm. For more information about our practices and experience, please visit bakerlaw.com.