Effective AI Depends on an Effective Data Ecosystem:

Megatrends Video Series

Tom [0:08] Welcome, and thanks for joining us. My guest today are Brian Jones and CJ Donnelly. They're both partners at guidehouse. It's good to have you with us.

Brian Jones - Guidehouse [0:16] Thank you, Tom.

CJ Donnelly - Guidehouse [0:17] Thanks, glad to be here.

Tom [0:18] And you know, almost no discussion about anything contemporary in information technology, whether it's networks, applications, modernizations, artificial intelligence, chat, GPT, all of these things, it always comes down to the data. And you have developed a concept that you call the data ecosystem that can help agencies with everything else they've got to do with data. We'll get into that. But let's start with what exactly do you mean by data ecosystem?

Brian Jones - Guidehouse [0:46] Well, thank you very much, Tom. So we are both members of a data ecosystem, and also consumers of data. So even on my way to the studio, today, I use a GPS application, I use the heartrate monitor, and we are generating gigabytes of data per person per year, as members of a data ecosystem. We're also consumers of a data ecosystem, and we consume data in our personal lives. So I like to wake up in the morning and really see, you know, what were the scores of the sports teams that I'm following? What's the weather gonna be like? What's my traffic going to be like? So you really start to create that data ecosystem around your lives as we consume that data. And data has progressed pretty fast in the last eight months to a year with the onset of capabilities like chat, GPT, and large language models. And there's been an increasingly disconnect between how we use data in our personal lives, and how we use data in our work lives, both the government and private sector. So as a physician, and prior federal employee, I firsthand saw some of that disconnect between that optimized data life I'm using in my personal life. And really some of the capabilities I'm actually using as part of

Tom [1:56] my work life. So now you're a data doctor, as opposed to a physician. Exactly. CJ, your thoughts on the data ecosystem?

CJ Donnelly - Guidehouse [2:04] Yeah, I mean, you know, just to tack on to what Brian was saying, the amount of data that's generated is unbelievable. You know, even just from this conversation that we're having, right, the three of us talking for 20 minutes, there's going to be a video that has video and audio, this is going to be posted as a podcast as audio only, maybe there's a transcript that's created, that's text only for the closed captioning, you know, maybe guidehouse decides to clip this off and post it on Twitter, and on Instagram, and LinkedIn. And just from the short conversation with three people, we may generate hundreds of different types of data structured and unstructured, complex and long and short.

And and it's just from the three of us from a short conversation. And imagine the amount of data that's generated in federal agencies where you have 1000s of employees that are interacting with at times 10s of 1000s, or hundreds of 1000s of customers on a regular basis. It's just incredible the amount of data that's generated. And so you know, we talked about the data ecosystem, and what you need to do to manage and really maximize all that data that's being created.

Tom [3:13] And you bring up a good point, especially in the last few months, eight months, a year or so, you have not only generating large data from applications and the general work people do, but by feeding them into some of these AI applications. The data combined generates yet more data. So you have data, creating data, which I think is a little bit of a shift from the way things were thought of prior to the advent of these particular programs. Fair to say, yeah,

Brian Jones - Guidehouse [3:40] it's that shift has definitely been part of the leg and that disconnect between how we use data in our personal lives and our work lives. So I think from a professional perspective, both from a government and the private sector, we've had to think about the rules beyond that data governance, should we be using certain types of data? So as a physician, you know, should we be using somebody's financial data with the medical information? So traditionally, there's been some firewalls both from a technological perspective, and from a business perspective in between those data models. But what we've seen with these types of capabilities is our ability to bring in different types of data. And both government and industry are really thinking about how do we put in rules in place to be able to maintain privacy, but also leverage some of the exciting capabilities that these types of ecosystems present.

[Tom 4:29] And I guess the question then is, you don't want to have data as a large, undifferentiated blob, I mean, ecosystem has system in the Word. And so not just the generation of data. So how do you enhance your data, including the system you have to store it, ship it, support it, protect it, and so forth? So that it becomes useful even as it keeps growing?

CJ Donnelly - Guidehouse [4:51] Yeah. And you know, I like to think of it as almost like a natural ecosystem, right. And let's take for example, you know, you're going out on a walk in, you see this like beautiful flower, right? You know, I think of that beautiful fully blooming flower as this, you know, advanced analytics visualization that providing analysis that you can make decisions on right. And everyone wants that, that beautiful flower. But but not realizing that you also have the soil and the sunlight in the rain and the insects that interact with it and spread the the pollen and the cross pollination and all these different things that interact in order to get that beautiful flower. And if you're missing any of those pieces, you don't have enough sunlight or you have too much water, there aren't the insects and you're not going to get the same type of flower, you may not get a flower at all. And so I think that's the the most important thing for federal agencies to think about is looking at the entire data lifecycle, together and holistically and not independently in silos. So you have the data management and the data strategy, combined with the the ETL, and the data engineering, to provide data for the data scientists to do the analysis. And then you get the visualizations and the reporting. All of those are interconnected, you can't do those independently. And so that's why I think like the advent of the chief data officer, as a rule in the federal government, has been so important, because now you have a senior executive who's responsible for that entire lifecycle and can be involved in that entire lifecycle to make sure that that agency that federal agencies are maximizing their data, so they can make useful decisions from this, trillions of records of data that's being created on a daily basis. And

Tom [6:32] that brings up the question of how do you make sure that you're using the right data, because sometimes, for a given application or a desired outcome, too much data, could be too much data and you wouldn't get what you want. So there's has to be some selectivity. Just don't dump all your data and every algorithm that you ask a question of fair to say, oh,

CJ Donnelly - Guidehouse [6:52] that's absolutely fair to say. And that's why it's really important to have data scientists and data engineers that understand that right, going back to the earlier example, about all the information that we're creating here, we have video clips, right? We may create 15 different video clips and the audio files and the video files and the transcripts and all of that, well, do we need to store all of those? Or could you just store the one full video clip and extract the information as you need it. And so just because you have more data, doesn't mean it's better data, or it's useful, and being able to look through the data and understand what's important, what's valuable, what's not valuable, what you need to store, what you need to clean, is becoming more and more important every day. Because if you just try to ingest all the data that's out there, it's too much, it's not valuable, it confuses the system, and you're not going to get the value in the results that you're looking for.

Tom [7:43] And in the generative field, you could even get results that are going in the wrong direction altogether.

CJ Donnelly - Guidehouse [7:48] Oh, yeah. Especially once you start training generative models on on data that's been generated from from the models and becomes this this cycle that it's generating its own data and what is real? And where's the analysis? And, and how can you determine what's effective. And so, as these large language models come into production, and everyone starts using them, which which people are more and more often, and government agencies are starting to explore how to use LLM to help their business is thinking about those considerations. And what are these models creating? What data should they be using? What data are they allowed to be using? And how can you take all that information together so that you have a valuable large language model and not one that just you know, hallucinating? or providing results that are not accurate at times can be dangerous. It's

Tom [8:34] almost like a ship that leaves the harbor, if it's one minute off, it will get through the bullies. But it'll Miss Hawaii by 1000 miles,

Brian Jones - Guidehouse [8:42] exactly they're applying those rules in place really are critical to leveraging the power in these capabilities. So if you think about just from your visits to a doctor, back in the day used to walk into a doctor's office and all your information set a manila folder on their desk. Now we will transition to electronic health record systems where now all of its in in the electronic health record system. Now, as we start to think about really leveraging the ecosystem of data from devices, smartwatches data that's collecting medical information, how do we start to bring capabilities together to start to bring all that to a workflow to a coalition because that now becomes part of the decision making process. But the rules are in place to really have some guidance around too much information is bad. But how do you get to the right information at the right time in that in that workflow, regardless of what your workflow is? Is that key question that really harnessing the powers of these technologies and the data ecosystem has to offer us

Tom [9:39] and you made a comment before Brian about the fact that there's a lag between data in the use of our personal life use versus in our professional or working life use. Can you elaborate on that?

Brian Jones - Guidehouse [9:51] Absolutely. So you know, what we said before really thinking through, you know, we have our lives optimized around data, so I know you know The system my apps know what sports teams I follow. They know geographically what whether I'm kind of interested in. And really, if we've got, you know, shopping apps know what I want to buy, I think before we want to buy them. But as we start to move to the federal space and the in the commercial space, you know, we are thinking through Hey, what are the legal frameworks for some of this data? Is data copyrighted? Am I allowed to use this type of data? Is this data for sale, we're not giving the right to that. So these are all aspects of data that the technology has progressed, sometimes faster than our federal legislation and our business rules or legal rules around that use of the data.

So that leg actually is kind of intentional, as we start to figure out what are those second and third order impacts of bringing in data that traditionally has been not integrated together, and actually allowing different parts of our federal and commercial health systems and the government, they use these things and we have to work the business model around that that's part of contributing to that lay. Okay,

Tom [11:00] we're gonna get into this a little bit more right now we're going to take a short break. My guests today are Brian Jones and CJ Donnelly, both partners at guidehouse. I'm Tom teman. This is the evolving complexity series data ecosystems, sponsored by guide house here on Federal News Network. Welcome back to the evolving complexity series, data ecosystems sponsored by guidehouse. Here on Federal News Network. My guest today are Brian Jones, and CJ Donnelly, both partners at guide house. And I'm Tom Teman, host to the federal drive. And let's talk about this idea of enterprise intelligence and kind of an emerging concept. They used to call it maybe knowledge management or something. But somehow, the growth in data that we've discussed has shifted the whole thing in a step function kind of way. So what is enterprise intelligence? And how can it use? How can it be used to help employees, their experience and their efficiency? Yeah,

CJ Donnelly - Guidehouse [11:55] absolutely. I mean, so all this data that's been generated and collected and analyzed from an agency or from an organization, you know, creates a sort of intelligence that exists within the organization that only that specific agency organization has access to that specific type of data, and the value and insights that can be drawn from it? And how can specific organizations use that to the benefit of their employees, right. And one of the things that we look at is this, this technology called like case complexity. And it's a way to help with caseload management that combines both automation and artificial intelligence and machine learning, to one better manage the case loads for workers. And so let me give you an example. Almost all federal agencies have some type of security clearance process that you have to go through whether it's a public trust or suitability, or whether you have to get a secret or top secret clearance. And there are standard forms that are completed SF 85, and SF 86. Is that are, you know, hundreds of pages long, and some can be incredibly complex and, and take, you know, months, if not years to be fully adjudicated. And we've used AI ml, and automation through robotic process automation, to be able to determine, you know, estimate, how long is it going to take for this, you know, individuals security clearance process to take, which ones are more complex may may involve a more senior adjudicator or may be able to distribute that workload to adjudicators who have more time to better balance the workload, also, at the same time, removing a lot of the manual effort that had to be done by transposing information from one system to another system, reducing the amount of time that an individual and adjudicator has to spend on that. So the end result is adjudications happen faster, they're better, they're more accurate, and the individuals doing them have a better experience doing them because they've eliminated a lot of the manual and tedious work they had to do. And so it's a win win for everyone, the agency, the individuals, doing the adjudications and the people that are going through the process, they can have a more expedited process.

Tom [14:06] So that's highly related to that idea of improved workflow, because it sounds like this optimizes the workflow, such that you can get more efficiency and throughput of in this case, case management.

Brian Jones - Guidehouse [14:18] Absolutely. So how do you begin, we talked a little bit the first segment around bringing data together from the ecosystem. And with CJ talks about, you know, artificial intelligence and large language models. What that's really doing is bringing the experience the knowledge of our workforce together over that data. So really, what we're doing is creating almost like a driver assist over the data ecosystem. So I am going to date myself a little bit, I learned how to drive a car that didn't have a rear view camera on it. And getting a rear view camera on a car actually is a new data feed. So I had to figure out am I backing up? When do I use the camera and when do I continue to look in the mirrors? So what we're doing is really layering that experience on to beginning to build out Got a virtual or automated Driver Assist at all of us and how we do our jobs, which is really exciting. Yeah,

Tom [15:05] I'm chuckling a little bit, because when I learned to drive, we didn't have an automatic transmission. Backup Camera really seems like a George Jetson space age type of thing. But this idea of optimizing, I think, is really what is at the heart of modernization in many ways. It's not so much which processor and server or whether you have a cloud, but it's really the experience, both for the constituent that is the reason it's going to wait for that Sf 86 to be processed. And also the adjudicators who face a mountain of work, they sometimes wonder if they'll ever get out from under.

CJ Donnelly - Guidehouse [15:39] Yeah, and you know, when we talk about optimization, and workflow balance, and it's really about making things better for the employee better user experience, both for the people who are doing the work, and for the people that are affected by the work. It's not about you know, how much information can you process? How much data can you ingest? But how can you take that information? How can you take the tools, the technology, and the people with the skills to actually improve a process to make things better for the for the federal agencies to respond to their missions better? And so it's not about the amount. It's not about, you know, the technology, but it's about how you can combine all of these things to maximize the data and improve the day to day for all parties involved?

Tom [16:25] And I want to get back to the generative question, because that is on on a lot of agency minds. And I just did my first generative search on one of the search engines and said, Would you like to try our generative? And it was a topic so narrow, that I knew the results made sense. But it wasn't one of these big, you know, what is the meaning of life questions where you could get anything to come up? But maybe comment, Brian, if you would, on, the best way to use the generative branch of AI, in a regulated controlled, you know, federal setting. So I think

Brian Jones - Guidehouse [16:57] most organizations are taking kind of a crawl, walk run approach to generate AI. So a crawl really is looking at how do we bring the data together and understand what data these these programs are reaching into and looking at, and really kind of making sure that those data sources are credible. So even at a generative AI search, you can specify I only want it searching these credible news sources, these credible source data sources, then as you start to actually progress a little bit, you know, putting in more of these quantifiers really understanding, you know, having a subject matter expertise around what you're asking for and being able to vet, you know, the answers you're getting back, there's where we stand today, there's still an opportunity for a person really to sit in the middle of data, generative AI searches, and truly operationalizing that data, it really gets down to you know, one of the key factors in integrating large language models is trust. So trusting where the data is coming from, and trusting those algorithms behind the surge is to give you the information that's that you can then truly act on.

Tom [18:05] Let me ask you this, in your experience, what could be the implications for medicine, because in one form or another, from research to clinical delivery, the government is the largest healthcare organization there is.

Brian Jones - Guidehouse [18:19] So the evolution of medicine has been interesting. So we went from the entire paper record to when I tried electronic health record systems, both on the federal side and the commercial side. Now, it could be argued that the electronic health record system created more work for doctors and nurses and technicians across the hospitals. And the the excitement around large language models and AI is really starting to make the jump that now we have the patient records digitized, we now have the opportunity to layer this technology on to make our lives more efficient to make them a little bit easier. So what it really does is allow us you know, as a clinician, I'm really excited to start to see the 10,000 medical studies that happen every year, that evidence in context with the person, the patient that I'm seeing in exam room. So we're starting to see bringing together current data, current evidence with the electronic health record. So bringing it together and making patient care, improving the quality of patient care, and also improving the experience of staff across our federal and commercial organizations. By integrating this technology into the workflow, it strikes

Tom [19:27] me that government could really be a good arbiter of the data of quality for this or any other application, because let's face it, there's a lot of misinformation on medical topics out there on the internet. And so you would not want those ingested into any kind of generative clinical setting,

Brian Jones - Guidehouse [19:44] right. And the government is really one of the perfect examples because if you think about the VA and the DoD put out clinical practice guidelines for those top conditions that you know are seen in both of those health systems. And we're talking about trust. Those are trusted sources of data. data that we're able to use to provide care across the federal organization. So it is really that the best petri dish to start to play to start to integrate generative AI technology into clinical care.

Tom [20:12] Yeah, so then data management and data curation, almost are important as they are maybe a little bit less important or subordinate to data stewardship and provenance control. Yeah,

CJ Donnelly - Guidehouse 20:27 it's incredibly important, right? Because you know, with all the data that's out there, understanding who should have access, who shouldn't have access, who can generate, who can't generate, who can edit, there's so many things to think about when it comes to data, and data sharing, that are incredibly complex. And if you get it wrong, you can have disastrous consequences. And I think that's when we talk about large language models and individuals versus organizations. You know, it's easy for an individual to opt in and say, Yeah, I'll share my data I'm comfortable with that. I'm comfortable with my searches, you know, being used in the model. But from an organizational wide perspective, it's a much different decision on what information is shared. How is that being used? Where's that data being stored, you know, with the with the federal government is very important that this information is not stored offshore? And and who's using it and, and who's taking the information and using it a way that maybe we're not comfortable with? And so those are the when you talk about data stewardship and data privacy, data protection, all incredibly important to the federal government. All right,

Tom [21:29] and the minute we have left, I'll give you each 30 seconds. What's the most exciting thing you're looking forward to? And what should agencies care about the most in the next say, year when it comes to this whole data ecosystem? CJ? Yeah,

CJ Donnelly - Guidehouse [21:40] I mean, I think it's about leveraging this, this emerging technology, and really getting the maximum value out of their data us that agencies can make better decisions with all this data that's been generated. And final

Tom [21:53] word, Brian,

Brian Jones - Guidehouse [21:54] I'm really excited about taking the data and the knowledge that is across the organization. Because when we're working, whether in the government or in the private sector, we're limited by what we know and what the others around us know. Now they begin to start to harness the power and the knowledge and experience of our entire workforce, around an increasing body of data to make all of our lives, you know, essentially more efficient, easier as we go forward.

Tom [22:19] Okay, well, well stated. Thank you both for being with us today.

Unknown Speaker [22:22] Thank you very much, Tom. Thank you. Yes,

Tom [22:24] two bins Brian Jones and CJ Donnelly. They are both partners at guidehouse. I'm Tom Tim and you're listening to Federal News Network. For more on this discussion, please visit Federal News network.com and search guidehouse

Transcribed by https://otter.ai

About Guidehouse

Guidehouse is a leading global provider of consulting services to the public sector and commercial markets, with broad capabilities in management, technology, and risk consulting. By combining our public and private sector expertise, we help clients address their most complex challenges and navigate significant regulatory pressures focusing on transformational change, business resiliency, and technology-driven innovation. Across a range of advisory, consulting, outsourcing, and digital services, we create scalable, innovative solutions that help our clients outwit complexity and position them for future growth and success. The company has more than 17,000 professionals in over 55 locations globally. Guidehouse is led by seasoned professionals with proven and diverse expertise in traditional and emerging technologies, markets, and agenda-setting issues driving national and global economies. For more information, please visit guidehouse.com.

🌐 guidehouse.com

✕ @guidehouse   linkedin.com/company/guidehouse

outwit complexity™