**RedZone Podcast Episode #45: AI and the Intersection of Psychology and Human Behavior – with Daniel McDuff**

Bill: I want to welcome you to the show today, Daniel.

Daniel: Thank you very much. It’s great to be here.

Bill: I’m very excited to have you on the show today because I think my listeners are really going to learn a ton. I love your work, your research. As I understand it, it is the intersection of psychology and computer science. I know that you’ve been at work really understanding human behavior. I’m really fascinated in understanding what is the push to have computers understand human emotion from your perspective?

Daniel:

[00:01:00] There’s several different levels to this. The first being that in order to understand human behavior and really drive insights in psychology to know more about how we behave, how we interact with each other, how humans communicate, we need to be able to measure behaviors on a large scale. That’s something that hasn’t been possible before. Building computers that can capture these signals, nonverbal signals, things like facial expressions, voice tone, even people’s physiology allows us to collect data from much larger scales than was possible before, where in the past people had to manually code data. If a computer can do it, we can do it much more efficiently.

[00:02:00] That’s really important if we want to build models of how people’s expressions vary across coaches, across gender, age groups. Then, the second really natural extension of that is if computers can measure these signals that humans are communicating through their behavior, then computers can naturally respond so that if the interfaces that we interact with, whether it’s a cell phone, whether it’s a TV or new [inaudible 00:02:04] headsets and the games that we play on those, if they can naturally respond to our emotional state it will make those interactions much more seamless. They’ll blend into the background of our lives rather than technology being clunky devices that seem to ignore our emotional state. They will appear much more human.

Bill: I understand you need large data sets to do that. Is that just because of the statistics around or the cultural expressions vary across? It varies across cultures and also across genders?

Daniel:

[00:03:00] Exactly. There are many factors which influence how people express themselves. There’s large amounts of individual difference. In order to get statistically significant patterns within the data, we need to measure it on a large scale. There are many parameters as you mentioned. Things like culture, social norms that exist in different countries influence how people of different genders express themselves. Situations influence what we express because in some cases it’s more acceptable to smile. In other cases, it’s less acceptable to smile, for example. Context is a huge factor to understanding emotions and being able to interpret the signals accurately.

Bill: I was reading about some technology that you developed. Maybe you could just elaborate on this that it was used in the Super Bowl. It was somehow related to measuring emotion at the Super Bowl and interpreted in the cloud. Could you expand on that just so that my listeners can understand that first for a moment?

Daniel:

[00:04:00] Yeah. That’s one of the first experiments we ran on a large scale trying to measure people’s emotions. One of the big challenges of measuring people’s emotions is how you create a situation where people are likely to express how they feel and be willing to opt in and contribute that data. It was at the beginning part of the year. This was back in 2011. We realized that people love Super Bowl ads. People sometimes just watch the Super Bowl for the ads. They’re very entertaining. They elicit a lot of emotion. We decided to create a website where people could choose to switch on their camera while they watched these ads and we’d measure their emotional response.

[00:05:00] We didn’t pay people to do this. Actually, we just put the website online. We advertised it through social media and through a few different press websites. People came to the site and they chose to switch on their webcam and we collected thousands and thousands of reactions of people in natural settings expressing themselves while they watched these ads. That was the first time ever that anyone has been able to in a sense crowdsource emotion data to start to build models that really feature a significant sample of more than say 100 people.

Bill:

[00:06:00] That’s really interesting. I was just listening to a show the other day about the difference between EQ and IQ and the emotional intelligence versus this high intellect IQ, the standard measure of intelligence. I find it very, very interesting that you are really in both ends here trying to marry both of these worlds. I’m not sure if you’re trying to marry them together, but how did this all develop your passion for this? It seemed to be almost at odds at some level.

Bill:

[00:07:00] Yeah. They do I think on the surface, but actually, true intelligence contains parts of both. We need both rational processing, but also, actually our emotions help us make better decisions in some cases. There’s a lot of research around decision making and emotions. Actually, when we have to make split second decisions our emotions play a part in helping us to manage those and make choices. We don’t just operate as purely rational decision makers. Actually, it’s a good thing that we don’t. That was one of the motivations for doing this research was to say, “Well, if that’s how we act as humans, then maybe computers understanding our emotions could help us make even better decisions,” rather than just ignoring how we feel and ignoring that we’re frustrated or angry.

[00:08:00] We can all associate with the fact that we’re in a bad mood sometimes. We make different decisions than we would if we were in a good mood. A computer being able to understand that could help us in our decision making, but then also there’s many other applications beyond that. Understanding and being able to track people’s health has huge potential. Emotions play a big role in how mental health is and how we can also recover from many different illnesses. Having technology that can help us monitor that and track it over time would be really useful.

Bill: It’s interesting. I was reading about your technology that you had developed or that became part of the cardio. I’ll link to this in the show notes, the cardio application for heart rate. I thought this was fascinating because I have been frustrated in trying to find an application that would be able to measure my heart rate without having to wear a strap, and I didn’t really understand what was so challenging about doing that. Maybe you could explain how you’re using remote techniques to measure heart rate.

Daniel:

[00:09:00] Yeah. This originally came about when we were looking at videos back in the Media Lab MIT. Part of the research in the Affective Computing Group was looking at facial expressions. Ming-Zher Poh, who’s the lead on the Cardiocam project, identified that you could use a camera signal or a video of a human face to capture the heart rate of an individual by doing video processing and looking at the amount of light that’s reflected from the skin and how that changes over time. We built the first version of that algorithm in the media lab.

[00:10:00] Then, that spun out into an iPhone app, Cardio, which allows people to do this just from their cell phone. This is really exciting because now we can analyze a video and not just extract what people are expressing on the outside what they’re facial expressions are, but also understand how they their physiology is changing. That gives us a measure of someone’s arousal, perhaps, or how excited they are even if they’re not expressing something on their face. It’s another dimension to someone’s emotional response.

Bill: Yeah. It seems that has a huge impact globally with the next billion that come onto the internet over the next couple of years. Just everybody. I forget what the percentages are, but vast numbers of them have cell phones. Then, that would be an interesting way to democratize healthcare across the planet potentially with being able to remotely take someone’s pulse, which would be interesting to see what the health implications of that are long-term.

Daniel:

[00:11:00] Exactly. One of the most exciting things about that project was it didn’t require any other hardware than is available on almost every consumer electronic device that we have now. Even our TVs have cameras, our laptops have cameras, our cell phones of course have cameras. Even new smartwatches will have cameras. Just by doing signal processing of the video stream, we can capture a signal which otherwise would require a customized device to be able to capture. That was one of the most exciting things. Of course, because it’s done from a video if you have multiple people in the scene, you can also measure the heart rate of everyone at the same time. You’ll need one device to capture the heart rate of say three or four people, that whereas in the past you’d have to give everyone a wearable in order to capture them.

Bill:

[00:12:00] Wow, that’s very interesting. Yeah. The health implications are very big. What about education? I know that’s one of your big interest areas. I could just imagine that if my son were taking a math test online or doing something like that and he was frustrated, I’m imagining that emotion. If the intelligence were sufficient of the math testing algorithms and it could sense his emotion, it might be able to adjust the teaching. Maybe you can tell me where teaching is going in education as it relates to this topic.

Daniel:

[00:13:00] We’ve seen in the last few years that online learning is exploding with many university courses now being offered online. Lectures can be viewed through platforms like edX and Coursera. That’s really popular because it’s democratizing the access to high quality education, not just fundamental material. People can now take exams online. They can get qualifications of a kind, but in the process of doing that, in the process of allowing large amounts of people to access this content, instructors, educators have lost a lot of the nonverbal cues that they used to get in the classroom.

A teacher in a classroom in the past may have been able to recognize when people were confused by what they were saying or when they were really engaging with an idea. Now, a lot of that information is lost because people are remote. They’re watching these lectures asynchronously. If cameras could capture some of their emotional response, a cognitive state while they engage with the content, then that could be aggregated and it could help educators improve the quality of their content. That, in turn, obviously would then benefit the students as well.

[00:14:00] In addition, a student being able to look at how their emotional state is changing over time while they engage with these courses may be useful. Sometimes when we get into a flow state which is a state of high levels of productivity, we don’t actually recognize what thing helped us reach that particular state. If we can have systems that log that type of information, we could potentially learn from it and be more productive overall.

Bill: Have you read the book by Steven Kotler, The Rise of Superman?

Daniel: I haven’t.

Bill:

[00:15:00] It’s about flow states. A light bulb just went off as you were talking because gosh, what a way to put that back from your education example. If you put that power back in the student and all of a sudden they saw their anxiety level scores rising as they’re taking the test, but if they were aware of that and if it was right front and center and they were given strategies to self-modulate, they almost could modulate their reaction to the material and put themselves back into that flow state. What a powerful education tool that would be.

Daniel:

[00:16:00] Yeah. I think that’s a great point. There’s a balance here. I think with this technology, it’s really important that the end users get access to all their data, and also have the ability to share it when they want to and also keep it private when they want to. Also, how we communicate the data to the end user is useful. Just being told that you’re getting more stressed may not be a very helpful thing, but being told ways that you could manage that and help you regulate it could be really useful. There’s definitely open questions about how emotion data is communicated to the end user and also how they can share that in order to make the most of it.

Bill: I’d love to get into some of the technology behind all of this. That may be one way we could start out with this is maybe could you define what the difference is between machine learning and artificial intelligence for our listeners? Don’t be afraid to be technical on the group here. I think it’s important for you to set a stage for what we’re talking about from a technology point of view.

Daniel:

[00:17:00] Yeah. That’s a great question. Actually, the line between machine learning and AI has become rather blurred, especially in the popular press over recent years, especially with the explosion of things like deep learning and platforms for recognizing voice and images very precisely. Machine learning is a very broad term which covers all sorts of things related to teaching a computer to recognize patterns, essentially. If I give a computer an image and I trained it to recognize the difference between cats and dogs, hopefully it will be able to extract the relevant features from that image and determine whether it’s a picture of a cat or a dog or neither. AI is a bit more philosophical.

[00:18:00] It’s about what it means to create intelligence. It’s about does a system need to be creative in order to be intelligent? Not just recognize the difference between two different things, but understand what that means. It can get a little as I say philosophical or complex, but really when we build machine learning algorithms the aim is eventually to build these skills or tools that an AI system will be able to leverage. A simple definition of AI would be to say, “Giving a machine skills that are usually attributed to a human.” In some senses, recognizing or interpreting emotions could be a form of AI.

Bill: Okay. Not for the complicated, recognizing emotions could be a form of AI because it would be an attributed human skill that now a machine has?

Daniel: Right.

Bill:

[00:19:00] Okay. Okay. That makes a lot of sense. Now, when I was learning about machine learning just a month or two ago, there was a difference made between deep learning and machine learning. I don’t know if you know Jeremy Howard. His company was analyzing. It’s called Enlitic. It was analyzing x-rays. They had taught the machine how to analyze x-rays better than a human being, a radiologist could analyze an x-ray. That was said to be more deep learning versus machine learning. Maybe you could just explain the difference between the two of those for our audience.

Daniel:

[00:20:00] Deep learning is really a subset of machine learning which is a broader term. In the past few years, there has been a lot of research, investment in techniques which allow us to build very complex or deep models using huge amounts of data. Typically, these models can learn representations of the data that aren’t prescribed by a human. In a typical or in a more traditional computer vision system, we would tell the computer to look for specific types of edges within an image or corners. We’d say, “These are the important parts. Now, extract those from every image and we’ll classify whether it’s a box or whether it’s a person or whether it’s a cow.” Deep learning techniques tend to work on the pixel level.

[00:21:00] That’s why they need a huge amount of data. Giving the computer this data, it then learns the representations that are important for discriminating between different classes of objects. The same applies to facial expressions or other types of computer vision tasks. The system has many layers. It learns representations on different scales. It was also using color and using edges and other types of information. It builds a very complex model which is typically the last stage of some form of neural network, which is the most popular right now for computer vision. It’s probably the convolution on neural network which is a system which does this feature learning to start with and then has a neural network at the end. Typically, these approaches perform really well on many different problems.

Bill: Is the work you’re doing a combination of deep learning because you’re at the pixel level and then also machine learning? Would that be a true statement?

Daniel:

[00:22:00] I would say in the past we’ve used a more traditional framework for doing computer vision analysis which was using, prescribing the features we wanted to be extracted, and then building a discriminative classifier on top of that. Now, we’re using deep learning techniques which work on the pixel level because we have enough data to really make good use of them. This is all still machine learning. Deep learning is really just a subset of machine learning.

Bill: Okay, perfect. That’s great. That gives a perfect context that I think people need to be able to understand. For lateral, where do you think … ? I know you mentioned the conversation. You used the word philosophy a couple of times. I find this interesting that philosophy majors as we have more exponential technologies like AI and machine learning, we’re having a corresponding rise in philosophy majors in universities.

[00:23:00] I have a question for you that I’m curious of your answer related to … We have this deep technology expertise that we’re building from machines to try to approximate human behaviors and human understanding. What about lateral thinking and drawing inference and the connections or connecting the dots that our human brain has? How are you seeing the intersection between your work and the more broader, philosophical connecting the dots type thinking that may be needed?

Daniel:

[00:24:00] Yeah. That’s a great question. In the context of emotion which is where a lot of my work is focused, to give an example, we can train a computer to recognize whether someone’s smiling, but that’s observing what they’re doing. If we really want to interpret the emotion behind that, we need a higher level of interpretation. What does this mean? Not just what are they doing, but what does it mean? That’s really the next element of building an AI system is that interpretation because that is something that is very complex. As I mentioned it before, context dependent.

[00:25:00] How you would interpret what the meaning is of someone smiling in a specific context may be different from another. What we’re doing now is to really integrate a lot of that context into the model. The culture someone’s from, what gender they are, what age group they are, what context they’re in, using all of that information to adapt these models to make better judgments of the emotional state of a person. Also, incorporating individual level baseline information and trends so that these models become more customized to individuals is an important part of that because there is a lot of individual difference in how we express ourselves and how we behave.

Bill:

[00:26:00] For an individual to benefit from this technology, I see several pieces of your work are really health related. Then, some are education related. Are there other areas of individual benefit that where an individual who owns a cell phone will be able to receive the benefits versus … ? I’m not saying this is bad, but on the advertiser side I can see where the advertisers can win if they can understand emotions with their advertising. On the human side, other than education and health, are there other areas that you’re looking at for potential wins from a general population point of view?

Daniel: Absolutely. Health and education are probably the two that I’m most excited about at the moment, but there are ways that we can make driving safer, for instance, using this technology, identifying when people are distracted and frustrated by experiences. That’s some of the work we’ve recently been doing is looking at driver behavior and trying to identify automatically when people are distracted or frustrated by an in-car system, some interface, whether it’s your navigation or your cell phone or something like that. Other areas of applications are in improving interfaces.

[00:27:00] I think we can all identify with cases where our computer has frustrated us when it keeps giving the same error message over and over again, even though we get more irate every time. I think we would all benefit. That’s maybe more of a less life or death type of situation, but I think we would all benefit from technology that’s a bit more sensitive to how we feel. That’s an example of another type of application. Our TVs responding more naturally to us, maybe our voice activated systems, whether it’s your Amazon device or your Siri. Actually, taking more care in interpreting your emotional state in addition to what you’re asking.

Bill:

[00:28:00] Yeah. I can see that with the Amazon Alexa or the Echo. Yeah. In Siri, I can see a lot of good application potential benefits there. Also, on the IT security side it would be interesting just from multi-factor authentication to different services to be able to use different models from both, not just your retinal scan, but facial. There was a gentleman I brought on earlier in 2015 who was using the second factor of authentication. He definitely was seeing that because we pushed so much processing power down to the phone, that we’re now going to be able to do a lot of the authentication functions at the phone because of processing power. There’s probably some direct application there as well. Daniel, how did you develop this passion? You’ve had some significant wins in your short career. I think you’re barely 30 right now. Right?

[00:29:00]

Daniel:

Almost.

Bill: Almost 30. This is a massive amount of success in a short amount of time. What do you attribute that to? What do you attribute your ability to take on these really cutting edge topics and really achieve a lot in a short amount of time?

Daniel:

[00:30:00] I had the privilege of being in a couple of different institutions. The MIT Media Lab, surrounded by just really fantastic colleagues and collaborators and also Affectiva where I direct the research. Rana el Kaliouby, the founder, has been a great mentor. I’ve had a lot of great collaborators who have been very creative. I think one of the things I’ve enjoyed the most and has been the most helpful is that in these environments it’s really been encouraged to try and if necessary, fail. In some cases, we can be scared to take risks or try to be ambitious with the projects we take on. Actually, it’s great to be in an environment where that’s not the case.

Bill: You’re in an environment where if necessary that you’re allowed to.

Daniel: Really, it’s been a product of the environment and some amazing collaborators.

Bill:

[00:31:00] That’s really interesting. Even though you’re in the East Coast, the last time I’ve really seen that culture was really in a Silicon Valley lab. It’s called the Hacker Dojo where literally one of the guys there saying that a failure is not a negative on their resume in the Silicon Valley area. It seems like the MIT Media Lab when you surround yourself with really strong collaborators and colleagues in a culture of taking risk and not being afraid to fail has really been a big benefit for you.

Daniel:

[00:32:00] Yeah. Another thing I would add to that is also sometimes some of the projects that seem a little frivolous or a little out there can actually be some of the most fruitful and most interesting. A lot of the projects that people work on at the media lab are related to their passions and hobbies outside of an academic context as well. By combining those different elements of life together, actually sometimes you get the most interesting results.

As I mentioned at the start, we started off measuring people’s emotions while they watched Super Bowl ads which seems like a fairly simplistic thing to do, things that make people emote. Actually, by demonstrating that we could capture people’s reactions online over the internet how they opted in with their cameras, it’s the first instance really of changing how we think about behavioral measurement, specifically of emotions and doing that on a large scale. I really do think that will change how we do observational research in the future.

[00:33:00]

Bill:

The observational research. Was that the passion in hobby you were referring to where you’re intersecting the work, the strong technology with the passion or was that the example you were referring to?

Daniel:

[00:34:00] Yeah. For me personally, I’ve always been interested in that human edge of technology. How we can interface it with … How we act in our daily lives. Personally, as I’ve mentioned that the healthcare applications are a big motivation for me. Thinking of ways that we can improve people’s quality of life and also help avoid some of the potentially tragic situations that occur, especially in first world countries, some of our biggest challenges around mental health and helping people to understand their emotions and how they feel. That’s really what got me motivated and interested in using the computer vision and machine learning knowledge to build things that help people.

Bill:

[00:35:00] Yeah. It seems from your work from the facial expression analytics and the wearable physiological measurement and analyzing large scale observational emotions, it seems there’s a real strong theme to your work that I think is going to have a really huge impact for people. I know a lot of my listeners are listening to it from what’s their impact personally within these technologies, but also within their own organizations? Maybe they’re trying to understand what’s coming down the line for capabilities. This is certainly the impact of your work as a broader human scale versus some of the guys and ladies on the phone listening. We’re going to figure out that the coffee machine is going to be smart enough to order its own coffee coming up right now.

It’s very different, but I think it’s important to understand where we’re going from that edge because we are going to have computer systems that are going to be mimicking human behavior. Obviously, that’s the good side. What are some of the things that we need to … ? This is back to the philosophy. How do we govern? How do we put a governing principle here? That while we have people charging down this path of doing a lot of good, that we also have some governance about what’s possible that can go wrong with this so that we have some governing frameworks in place. How do you address that from your research in the companies that you’re spinning out from that research?

[00:36:00]

Daniel:

Yeah. I think that’s a really important question. As an engineer, as a scientist, I take that social impact of the technology really seriously. I think again that’s something that is at the forefront of the research at the media lab too was really understanding what is the social impact of what you’re creating going to be? How can you measure that? How can you understand it? In a sense, how can you redesign things potentially to make the best impact possible?

[00:37:00] Building social norms around how people use this technology I think is really important, making sure that people have the opportunity to opt in when they want to, but also stay private when they want to. Also, when we do collect data as I mentioned before, making sure that the end user has access to that data. It’s not just being collected and sent to some organization. While that may be a good use of the data, the end user should always have the choice to access it if they want to, to have ownership over that data. I think that’s really important.

[00:38:00] In order for people to be able to really make the best choices about what they do with their data, we need to be able to communicate it to them in the best and most accurate way. With emotions that’s a tricky thing because we’re not very used to introspecting and analyzing our emotional states. Emotions are things we experience transiently, and then often we don’t reflect on how we felt two or three days ago. This technology now allows us to do that, but the computer needs to be able to communicate that type of data to individuals in a sensitive and helpful way. There’s some design aspects about that, that need to be [inaudible 00:38:08].

Bill: Yeah. It’s a powerful path you’re down. I know that Google and Facebook already do this from the text based what people are writing and the emotion around text. Obviously, this is a different level of magnitude of difficulty what you’re working on, but I can see it’s the next step. There’s a privacy impact that has to be managed as we go forward. It’s good to see that you’re actively working on that angle as well and aware of it.

Daniel:

[00:39:00] Yeah. I hope the debate is a very broad one. I think that we’re starting to see more of this conversation about personal data and privacy becoming more present. You think of the article only very recently about Apple unlocking a phone for the FBI. That’s an extreme example, perhaps, but I think it’s really important, that this is a very broad and public debate about how we design technology that both collects and stores and analyzes personal data.

Bill:

[00:40:00] Yeah, and the interesting piece. I’m not sure. I think context is going to be a big piece of this. It’s a big piece of security both corporately anyways is it’s not that someone accessed a system. It’s the context in which they did. Did they have the roles? Did they have the permissions? Are they coming in with a second factor or not? There’s so many. There’s a context to everybody that determines that expression. That’s an interesting piece I’m sure that what I’m understanding is you guys are still evolving context in the deepening of that as we go. On a scale of 0 to 10, 10 being perfect human being level context, where do you think we are at a 0 to 10 level of layering in context into understanding where you are at this point?

Daniel:

[00:41:00] In terms of recognizing emotions, machines are still very much in the infancy. Maybe a ballpark could be a computer is as good as maybe a four year old child at recognizing a facial expression and the emotion associated with that. There’s still a long way to go in terms of building systems which are really- That being said, even the skill of a younger child when that can be deployed at scale is still very powerful. You can think of all of the context where you could understand at least even with some level of depth at the emotions of a large group of people or people in different context or a single person over a long period of time and how useful that would be.

Bill: Daniel, I really appreciated our conversation today. Maybe just what would be the best way for people to learn about your work and/or the topics we’ve talked about that you think I could put on the show notes that people could learn more?

Daniel:

[00:42:00] Certainly, on my website there’s a lot of information about the different projects I’ve talked about today. The Cardiocam, measurement of emotions over the internet and the Super Bowl projects. By visiting my website, Daniel McDuff is probably sufficient on Google to find that, but also my web address. I’d be very happy for you to add that to the show notes. Also, more broadly speaking, the Affective Computing Group website from the MIT Media Lab is a great source of information about some of the latest projects in the affective computing space, the emotional computing space.

Bill: We had mentioned that. Just to briefly interrupt you, the affective group. Is that an offshoot of the MIT Media Lab?

Daniel:

[00:43:00] That’s one of the research groups within the media lab. Professor Rosalind Picard is the PI. She was my former advisor. Yes. That’s one of the research groups within the MIT Media Lab.

Bill: Excellent. I’ll definitely put notes out there for it. I just interrupted you and you were going through. You had just ended it [inaudible 00:43:12]. Is there any other places to go or is that a pretty good list right now?

Daniel: That’s a good list for now.

Bill: I definitely thank you so much for coming on the show, Daniel. This has been fascinating. I really appreciate you for the work that you’re doing in the world to really democratize healthcare and education and put these technologies in the world in a very powerful way.

Daniel: Thank you very much. It’s been a pleasure to talk to you.

Bill: Okay. We’ll look forward to talking to you maybe sometime next year.

Daniel: Great.