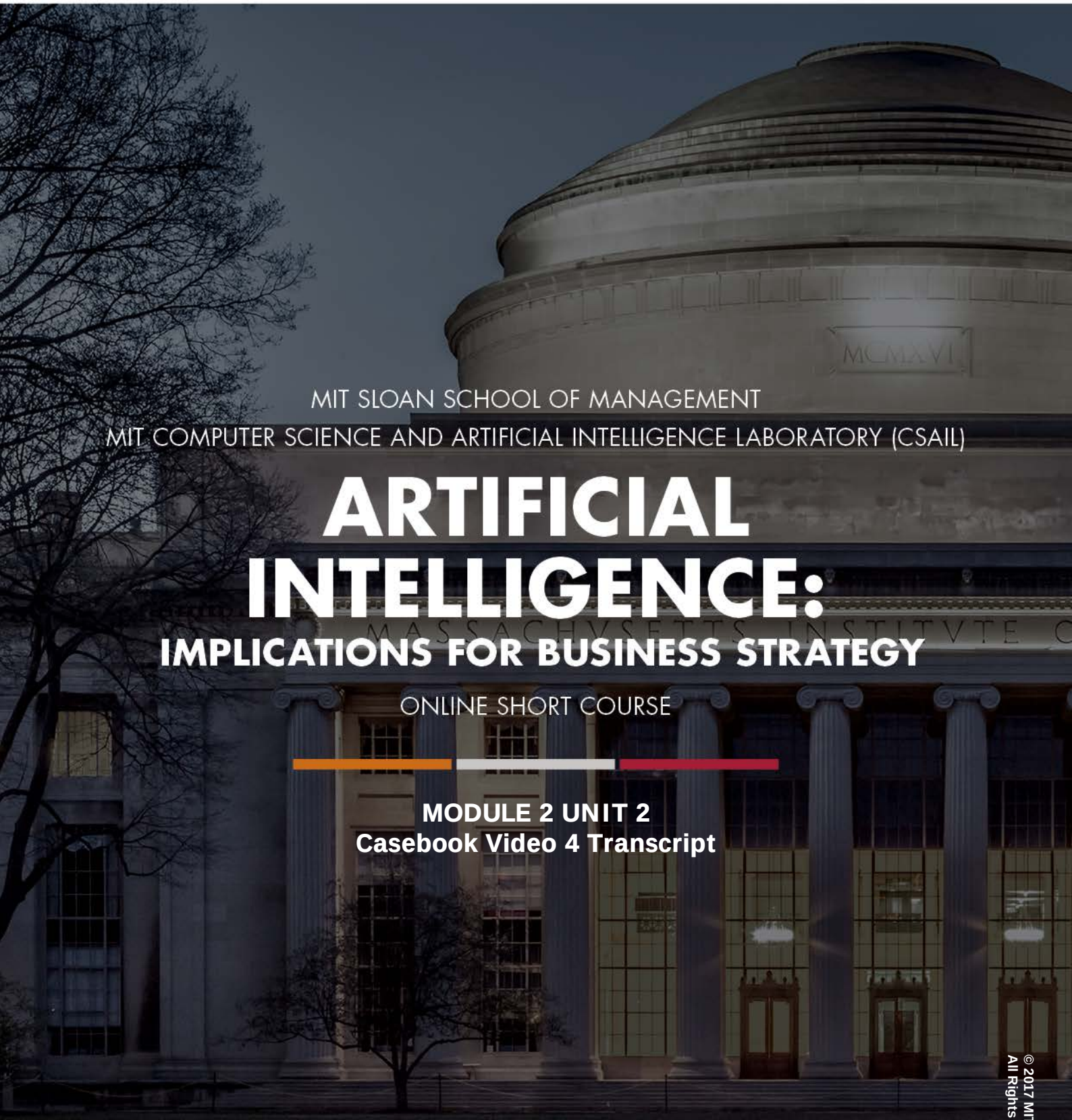
MIT MANAGEMENT EXECUTIVE EDUCATION

MIT CSAIL



MIT SLOAN SCHOOL OF MANAGEMENT

MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY (CSAIL)

# ARTIFICIAL INTELLIGENCE:

## IMPLICATIONS FOR BUSINESS STRATEGY

ONLINE SHORT COURSE

**MODULE 2 UNIT 2**
**Casebook Video 4 Transcript**



DELIVERED IN COLLABORATION WITH getsmarter

# Module 2 Unit 2 Casebook Video 4 Transcript

THOMAS MALONE: So, in the next module of this course we'll talk about natural language processing and why it's hard for computers, what computers can and cannot do with natural language. But in the next segment, which you're just about to hear, MIT Professor Alex Pentland is going to talk about something computers can do today to help people with natural language communication. In this case, however, he is going to talk about how to use machine learning to pay attention to the non-linguistic aspects of a conversation, things like the tone of voice and the timing, and to use that as a way of coaching call center representatives about how to have more effective conversations with the customers they're talking to on the phone.

ALEX PENTLAND: Hello, I'm Professor Alex Pentland here at MIT, and I want to tell you about honest signals and how that can be used to interpret human intention and emotion, and has been realized in a spinoff company, Cogito. So, about 15 years ago now I was setting up a laboratory in India and I had a board of directors full of superstars, and we were terrible. And the thing that was really bad was there was, sort of, like, too much charisma in the room. It was all these egos banging, and it wasn't the words they said, it was how they said it that seemed to be causing the problems. So, when I extracted myself from that situation I determined that what I was going to do is understand that unspoken language, the language that isn't words, it's how you say it.

And a way to think about this is that humans have had words only for a relatively short time, and before that we had signaling behavior. We had dominance and interest and things like that without words. Apes have that sort of signaling today, and importantly, we still have it along with the words. But usually we concentrate just on the words, even though that signaling behavior is so important in being taken seriously and dominance and expressions of interest and so forth. So, what I set out to do is I set out to begin to measure this using microphones and accelerometers and things that would capture the nonlinguistic components of human behavior, while people negotiated and gated and did other sorts of natural things that we humans are tending to do, and what I found was that there is indeed this sort of second language that people have, and that it's incredibly important in outcomes between people. In fact, it typically counts for maybe 40% to 50% of the outcome, and the words account for the rest, which is amazing because we always concentrate on the words and so rarely on the body language and signaling behavior that goes with it.

What this signaling behavior is, is it's the result of some of our basic neural processes. For instance, we have an autonomic nervous system, our fight or flight system, and this system gets us nervous energy when we're afraid or excited, and you can see that in how people behave. So, you think about a three year old who when they're excited they jump around and they won't shut up. Well, adults actually do much the same thing, and in fact we did an experiment with semi-pro poker players where they were trying to calm their excitement when they got a good hand, but they weren't doing a very good job, and so we could tell when even these pro players had a good hand, and this is probably the first social science experiment where the experimenter made money off of the subjects.

Everybody has that sort of signaling behavior, and it's hard for us to pay attention to it because we're so concentrated on the other things, the words for instance, but with computers analyzing the audio, analyzing the body language, you can tell when someone's excited. You can also tell when someone's paying attention, and you tell that by the timing between people. For instance, are there pauses where there shouldn't be pauses, do people almost cut each other off. It's almost like this dance between people in terms of the timing, and people who are paying attention, people who are really interested, have very tight timing. Again, as a human it's hard to actually measure this stuff and tell it, but as a computer you can do this really well.

So, things like your autonomic nervous system, your level of excitement, your mirroring system where you put all those little "uh-huh," "sure" those sorts of things into the conversation, those are things you can measure using the computer, and they predict how the conversation is going. So, as an example, is one of my spinoff companies, Cogito Corporation, has used that to analyze in real time conversations between call center representatives and clients. And so, what they do is they measure these signals, they use machine learning to know when things are going well and when they're not going well – they call it, sort of, collision avoidance – they provide these signals back to the call center person about, you know, when to slow down, when to be quiet, when to put something in, sort of helping them, coaching them in how to talk to people in a way that's going to be effective. And what they find is they find that by having this real-time machine learning on the nonlinguistic part, the tone of voice, the gesture, the voice gestures, they're able to increase customer engagement by often up to a third, be far more effective at conveying message across, and it's better for the call center representative also because they have less fights with clients on the telephone.

So, the turnover in the call center is much less, which leads to lower cost. So, you might ask, how does this technology actually work, where does the machine learning fit in? And so, what it does is it takes the audio stream between the call center representative and the client and it takes those two audio streams, it first extracts features that have to with activity level and timing, and then it uses machine learning to be able to interpret how well the conversation is going. And this is supervised machine learning originally, so we went through by hand and we annotated lots and lots and lots and lots of examples of conversations and how they were going, then we trained up the machine learning tool to be able to predict the hand labels from the features, and now we have something that's pretty robust for being able to take people with any accent – in fact, actually it can be in any language because it's not listening to the words – and figure out all these labels that have to do with how is the conversation going, is the person actually paying attention, etc., etc., etc.

So, the same type of thing can be used for diagnosis of manic depressive syndromes, and even for PTSD and depression, and of course we're working on that as a way of addressing these big health problems, not again by listening to the words but by listening to that nonlinguistic tone of voice and applying machine learning to it, to understand people's signaling behavior, as opposed to trying to understand their words.

THOMAS MALONE: Did you understand all the concepts covered in this video? If you'd like to go over any of the sections again, please click on the relevant button.