

Microsoft: The Dawn of the Cognitive Services Era

By: Rick Lievano

Every few years a new technology emerges to a point where it significantly impacts the world in which we live – both personally and professionally. In the last decade, the advent of virtualization and cloud computing significantly altered the software market, ushering in a range of new products and services and a new methodology for managing both. By contrast, “Big Data” has failed to reach the full potential that its proponents evangelized. Big Data has led to lots and lots of data, but only limited actionable results. But that’s all about to change.

Cognitive services, such as those in the Cortana Intelligence Suite, could be the missing link in resolving the Big Data revolution. Machine learning algorithms can provide value for collected data.

They can process the increasing petabytes of real-time streaming information. You will be able to provide automated decisions based upon a learned understanding of the data. Cognitive services are the next big-impact technology.

The combination of several mature technologies is now facilitating today’s break out of cognitive services. The convergence of four factors has triggered this: (1) the almost limitless computing capabilities of the cloud, (2) access to substantial data sets for training and analysis, (3) advances in algorithms and training methods such as deep neural network learning, and (4) the advent of commercially available SDKs and APIs that make connecting to these cloud services relatively simple.



It’s important to note that this isn’t brand new technology. As is the case with all emerging software revolutions, multiple players have been around a long time. We’ve been at it across various Microsoft product groups for years. Known in the past as Project Oxford, Microsoft researchers contributed to the advancement of the

science while building popular cognitive tools for customers. The Cortana personal assistant, integrated with Windows 10, has responded to over a billion questions asked by over a hundred million different customers. But it

was just the tip of the spear. Now Microsoft is bundling this experience into Microsoft Cognitive Services.

Cognitive Services, Defined

With the rapid advances in natural language processing and visual image processing, we have reached a nexus point in the evolution of this cognitive services. The most information rich senses for human interaction with the world can be recognized and analyzed by software. When these are considered together, complex machine



Not for distribution or reproduction.

behaviors can be realized. Microsoft has organized the Cognitive Services APIs around five domains: Speech, Language, Vision, Knowledge, and Search.

Speech

Billions of dollars and countless man years of research and development has led to several cognitive service skills being ready for prime time exploitation by any Microsoft customer. Speech Recognition has progressed to the point digital assistants and switchboard IVR replacements can understand realistic human speech. Applications using cognitive services understand naturally spoken language: context, meaning, slang, jargon, and even local accents. Microsoft researchers achieved the lowest word error rate of 5.9 percent when tested against the industry standard switchboard speech recognition task. This is human parity - the same level of understanding humans listening to the same conversation.

Language

Text can be converted to speech for automatic verbal responses. Speech can be transcribed as text and piped as input to the APIs. The Language Understanding Intelligent Service (LUIS) allows for models of specific populations and subject matters to be trained in context intent recognition. It can understand the relevance of the conversation to a specific subject domain, return the topic of the conversation, the key phrase in an exchange, and gauge the sentiment of the speaker.

Vision

Image Recognition can interpret real time images from multiple formats. An image can be recognized as a collection of objects and the actions.



In today's reality, cognitive services simply recognize emotions in humans to better understand the context of a situation.

Image Recognition can identify and assess the collective emotions and ages of people in an image. It can recognize the environment of the image and name the elements displayed. For example a landscape shot could return: "outdoor, nature, forest, tree, river, ravine, and gully." It can detect text in the image, such as a sign, and understands it. Image Captioning brings this all together. It allows a simple and complete story to be told about the image.

The Combined Power

The greatest potential of Cognitive Services is coming from the integration and combination of these APIs in novel mashups. A dramatic mashup in development at Microsoft is already easing the life of Microsoft's blind employees. The project Seeings AI. Computer vision and natural language processing are combined to describe a person's surroundings, read text out loud, answer spoken questions and identify facial emotions during conversations. For example the visually impaired can

take a picture with Pivothead smart glasses and the Microsoft CaptionBot will describe what it sees in the image, identifying objects, people, emotions, actions: "a happy

young girl is throwing an orange Frisbee.” Seeing AI can also be accessed from a cell phone. For example, it can give verbal instructions for positioning and snapping a picture of a document. It will recognize the document type as, for example, a menu. Then it can respond to commands to read aloud the appetizers.

Science Non-fiction

One of the biggest marketing obstacles of Cognitive Services is its big brother Artificial Intelligence (AI). The perceived endgame of AI research, creating actual machine intelligence, is a far-point on the distant horizon. AI, via Science Fiction, is seen as a future, not present, technology. Fictional AIs, like the transcendent companion of the picture Her, Samantha, or the diabolical nemesis of Ex Machina’s Ava, have emotions for dramatic tension. In today’s reality, cognitive services simply recognize emotions in humans to better understand the context of a situation. Alternative meaning in speech is often keyed to emotions. Accuracy transcribing speech and translating from culture to culture is aided by the correct interpretation of the emotional intent of the speaker.

Star Trek seeded many ideas for current innovations. One such was the 25th Century Star Trek communicator, a working replica of which you can buy today as a Bluetooth pin for your tunic. Coupled with this futuristic communicator was the Universal Translator, which after some deft training by the language savant Uhura, interacted with the communicator to translate between two different cultures in real time. Surprise, this capability is in



MEO, the television, Internet, telephone, and mobile quadruple play brand of Portugal Telecom, created a data-driven service to suggest which VOD content that a customer might wish to view.

use today. Your developers can access the APIs and, like Uhura, train a service to translate your business needs. Our mobile telecom partner Tele2 did just that.

The New Frontier of Cognitive Services

Europe is experiencing an influx of refugees and needs to accommodate these non-native speakers. Tele2 and Microsoft teamed up to address this challenge. They integrated the Translator API and the Bing Speech API into their network. This allows for real-time translation services in both their fixed and mobile phones. Tele2 is now reaching out to potential customers and are prepared to launch the service soon.

SoftBank and Microsoft collaborated with augmenting Pepper, a robot designed for human interactions, with these APIs. Pepper, in retail stores throughout Japan, will have Emotional Intelligence – understanding human sentiment and recognizing customers via facial recognition. Pepper will use cognitive services to correlate product and service recommendations to gender, context, cultural and situational context. From the customer’s perspective by understanding their emotions and reacting accordingly, Pepper will exhibit empathy. Microsoft Japan’s, Takuya Hirano explains, “Pepper will be able to recognize a customer who comes to the [store a] second time and to remember his purchase history... [and] offer a more personalized service. With the Microsoft Translator tool, Pepper will ... play the role of interpreter between foreign clients and [the] Japanese trader.”

This easy accessibility to Cognitive Services is new. Companies are just beginning to scratch the surface of the new technology. Nevertheless, dramatic business impacts can be achieved now. For example, with too many video on demand (VOD) choices, customers experience confusion and either delay their selection, or

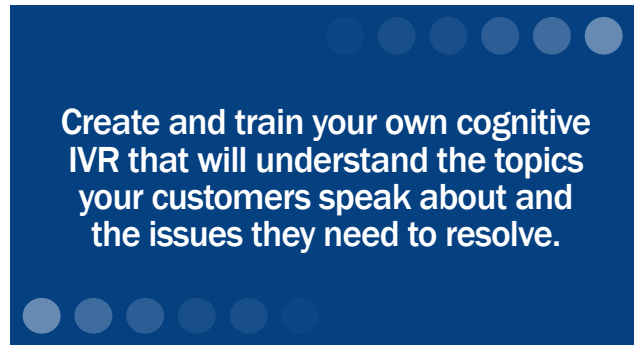
reject all choices. MEO, the television, Internet, telephone, and mobile quadruple play brand of Portugal Telecom, created a data-driven service to suggest which VOD content that a customer might wish to view. "The results are very impressive. The Microsoft Cognitive Services Recommendations API has helped us take a big step in improving our user experience and building more loyalty to our services," reports Jorge Rosa of MEO, Head of TV Strategy and Interactive Services. VOD consumption increased dramatically and customers began spending longer time engaging in VOD services. More premium content was purchased.

Quick Start

You can start leveraging Cognitive Services today. Take advantage of the power of data, the cloud and intelligence. Microsoft makes it easy to begin by making the APIs accessible here and by offering workshops to provide further assistance.

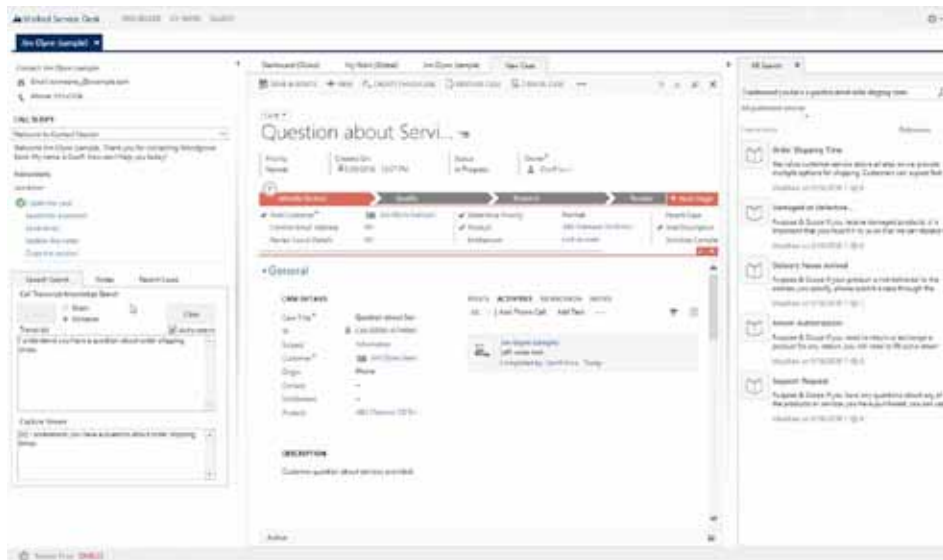
Microsoft created the Bot Framework to make it easier for developers to build and connect great Bots to users. The framework is integrated with the Cortana Intelligence Suite. Cortana is already trained with conversational styles that are culturally dependent. For example, for U.K. users who value self-deprecating humor, she uses playful sarcasm in her responses if she senses you're fooling around.

The payoffs can be immediate with small starter projects. Provide intelligent services to your contact center agents,



allowing them to better serve their customers. Envision a virtual agent assistant that helps the agent complete tasks faster, more accurately, and with better customer service. This Bot can act in the background on behalf of a human agent, translating and transcribing speech to text, then automatically fetching related knowledge.

IVRs, as they are in use today, are detrimental to customer satisfaction. Bots can provide direct personal customer care before a customer reaches an agent. In many cases, this can solve an issue without need of an expensive human contact center agent. Further, create and train your own cognitive IVR that will understand the topics your customers speak about and the issues your clients need to resolve. Automatically identify the language of the caller. Apply translations to over 30 languages. Fast track by first launching your Bot on Skype; the Bot framework is already integrated. Need to test drive a Bot concept? You can publish it to Microsoft's Bot Directory for crowd-sourced testing.



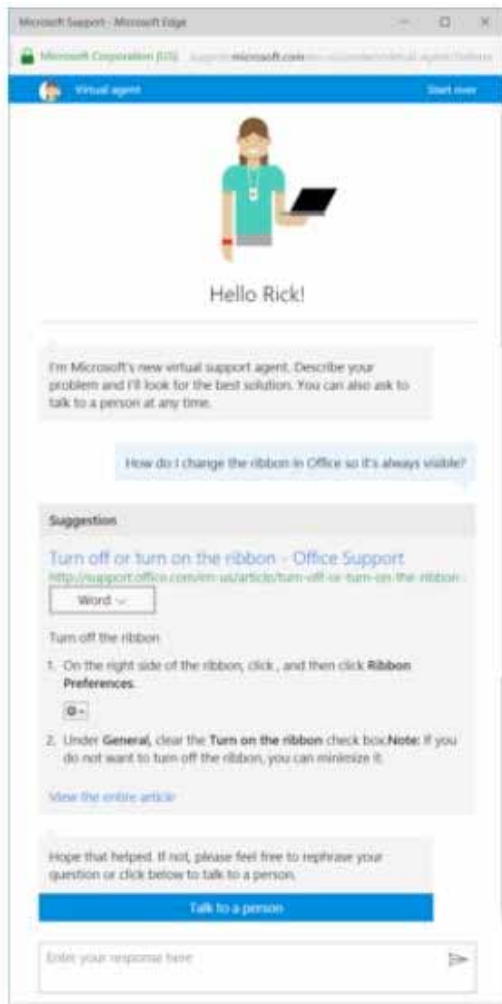
For example, you can access the Microsoft Support Bot here. "Get Started" will connect you with the virtual agent that will first try to resolve your issue.

For cognitive and data scientists, Microsoft provides the Computational Network Toolkit. The open source CNTK is designed to be flexible and easy to configure but it still outperforms other open source environments. Computation graphs are created from a high-level

description language and most training parameters are easily configurable. CNTK is what Microsoft's own engineers use in developing deep learning applications. The Cognitive Services Era

The world is about to change. With Democratizing AI, Microsoft plans to infuse every application that people interact with, on any device, at any point in time, with intelligence. These same intelligent capabilities that are infused in our own apps — the cognitive capabilities — will be available to every application developer in the world. To speed this along, we've just consolidated our AI teams, over 5000 researchers and engineers with the creation of the *Microsoft AI and Research Group*.

We've also entered into a groundbreaking, non-profit public partnership: co-chairing the Partnership on AI. Joining us are Amazon, DeepMind/Google, Facebook, and IBM. This is a new non-profit organization charged with



In the near future, mundane, routine tasks will be managed by your digital personal assistant and a plethora of friendly Bots.

identifying and addressing challenges and opportunities in an open and inclusive manner.

This is where the edge of time catches up to Science Fiction. Anyone can talk to any device. Microsoft calls this new paradigm, the Conversational User Interface, or CUI. In the past, we designed based on the input of mice and keyboards and the display characteristics of consoles. Now you can design around speech, snapshot images, and visual interpretations of real-world actions. Bots will independently trigger actions by interpreting real life circumstances. Cognitive agents can understand what is happening and make decisions. This goes beyond policy management. Cognitive Bots will take action triggered, not by policy you specified, but more significantly, the actions learned in the deep neural network.

In the near future, mundane, routine tasks will be managed by your digital personal assistant and a plethora of friendly Bots. Potentially any connected device, via commands sent into the cloud, could understand and follow your direction. This is the next era of communication technology, and it's easy to begin to tap the potential now. The pioneers will capitalize on the multitude of opportunities of this new frontier.