

Industry Watch

The return of the chatbots

ROBERT DALE

Arria NLG plc

e-mail: robert.dale@arria.com

(Received 10 August 2016)

Abstract

By all accounts, 2016 is the year of the chatbot. Some commentators take the view that chatbot technology will be so disruptive that it will eliminate the need for websites and apps. But chatbots have a long history. So what's new, and what's different this time? And is there an opportunity here to improve how our industry does technology transfer?

1 The year of interacting conversationally

This year's most hyped language technology is the intelligent virtual assistant. Whether you call these things digital assistants, conversational interfaces or just chatbots, the basic concept is the same: achieve some result by conversing with a machine in a dialogic fashion, using natural language.

Most visible at the forefront of the technology, we have the voice-driven digital assistants from the Big Four: Apple's Siri, Microsoft's Cortana, Amazon's Alexa and Google's new Assistant. Following up behind, we have many thousands of text-based chatbots that target specific functionalities, enabled by tools that let you build bots for a number of widely used messaging platforms.

Many see this technology as heralding a revolution in how we interact with devices, websites and apps. The *MIT Technology Review* lists conversational interfaces as one of the ten breakthrough technologies of 2016.¹ In January of this year, Uber's Chris Messina wrote an influential blog piece declaring 2016 the year of conversational commerce.² In March, Microsoft CEO Satya Nadella announced that chatbots were the next big thing, on a par with the graphical user interface, the web browser

¹ <https://www.technologyreview.com/s/600766/10-breakthrough-technologies-2016-conversational-interfaces>

² <https://medium.com/chris-messina/2016-will-be-the-year-of-conversational-commerce-1586e85e3991>

and the touch screen.³ And in April, Facebook's Mark Zuckerberg proclaimed that chatbots were the solution to the problem of app overload.⁴

Discounting for the usual proportion of hype that is bound up in statements like these, the underlying premise looks unarguable: interaction with technology using either natural language text or speech is becoming increasingly feasible, and potentially very significant. William Meisel, a respected commentator in the speech technology world, distinguishes 'general personal assistants' like Siri from the tsunami of more narrowly focused chatbots, which he calls 'specialized digital assistants'.⁵ He predicts that the latter category will generate global revenues of \$7.9 billion in 2016, rising to \$623 billion by 2020.

Worth a closer look, then.

2 The demographics of bot land

So what makes up this population of bots we can look forward to meeting on Digital Main Street?

For a start, we have the already-mentioned voice-driven digital assistants from the major players. Released in 2014 for Windows Phone, Microsoft's Cortana became available on the Windows 10 desktop operating system in early 2015; and in mid 2016, it was due to appear in the Xbox One interface. At the time of writing, Cortana is available in English, German, Italian, Spanish, French and Mandarin. Apple's Siri, which debuted on the iPhone in 2011, is now available on your desktop via macOS Sierra as of the middle of 2016. Siri adds a significant number of languages beyond those offered by Cortana. Amazon's Alexa, embodied in the Amazon Echo smart speaker, became widely available in the USA in mid 2015, but hasn't since learned any new languages, and you can't order it from outside the US. Google Assistant, announced in May 2016, is an extension of Google Now that is able to keep track of a conversation. Google Now supports a list of languages as long as your arm, but it's unclear how quickly each of these will acquire conversational capabilities.

All of these applications can help you with some subset of the standard virtual assistant skill portfolio, which generally includes scheduling meetings, checking your calendar and making appointments, reading, writing and sending emails, playing music, and, increasingly, controlling your suitably automation-enabled home. You'll find dozens of blog postings providing comparisons of the relative merits of the four assistants for these various tasks, although I assume the developers of each are avidly copying their competitors' best ideas and features, so any differences these bake-offs identify may be shortlived.

But these four ambassadors for conversational technology are really just the tip of the iceberg. The focus of Meisel's study, mentioned above, are digital assistants that operate in very specific domains or help with very specific tasks. Think of anything

³ <http://www.businessinsider.com.au/microsoft-to-announce-chatbots-2016-3>

⁴ <http://www.wsj.com/articles/facebook-hopes-chatbots-can-solve-app-overload-1460930220>

⁵ <http://tmaa.com/specializeddigitalassistantsandbots.html>

you might want to do on the web—book a flight, buy some shoes, take issue with a parking fine—and you can bet there'll already be a bot for that.⁶

Example: Just this morning, I got an email informing me of five new apps that have been added to the Skype Bot directory.⁷ The Skyscanner Bot lets you search for flights; the StubHub Bot helps you find tickets for events; the IFTTT Bot lets you build automated trigger-based messages from a wide variety of apps, devices and websites; the Hipmunk Bot provides travel advice; and if you're not all talked-out after that, you can even chat with a bot who's stolen the identity of Spock, second in command of the USS Enterprise, to learn about Vulcan logic.

There is already a vast bot development community. Pandorabots (www.pandorabots.com), which calls itself the world's leading chatbot platform, claims 225 thousand developers, 285 thousand chatbots created, and over three billion interactions. No doubt a very significant proportion of those apps are one-off experiments from tyre kickers, but those are pretty impressive numbers nonetheless. Since Facebook's April 2016 announcement of its tools for building bots that operate inside its Messenger platform, over 23,000 developers have signed up, and over 11,000 bots have been built.⁸ There are, Microsoft says, over 30,000 developers building bots on the Skype platform.⁹ Kik, an instant messaging app used by around forty per cent of teenagers in the US, claims that developers built 6,000 new bots for its platform in June alone.¹⁰ There's so much going on that you'd really want something like a magazine to keep you up-to-date with the latest happenings. Oh wait: there is one, sort of, at <https://chatbotsmagazine.com>.

And if you've cut your teeth with a simple bot for Facebook Messenger or Skype, and want to upgrade to play with the leaders of the pack, there are toolkits there too. The Alexa Skills Kit (<https://developer.amazon.com/alexa-skills-kit>) is a set of APIs and tools that let you add new skills to Alexa. Microsoft's Bot Framework (<https://dev.botframework.com>) works across a range of platforms. In June, Apple announced an SDK for integrating third party apps with Siri (<https://developer.apple.com/sirikit>). Google has its Voice Actions API (<https://developers.google.com/voice-actions>). To win in the conversational commerce land grab, you need to enlist an army of third party developers.

3 Hey bot, don't I know you from somewhere?

But let's back up a minute. For present purposes, we'll take the term 'chatbot' to refer to any software application that engages in a dialog with a human using natural language. The term is most often used in connection with applications that converse

⁶ <https://www.theguardian.com/technology/2016/jun/28/chatbot-ai-lawyer-donotpay-parking-tickets-london-new-york>

⁷ <http://blogs.skype.com/2016/08/03/skype-says-hello-to-new-bots>

⁸ <http://www.forbes.com/sites/kathleenchaykowski/2016/07/01/more-than-11000-bots-are-now-on-facebook-messenger>

⁹ <https://blogs.microsoft.com/blog/2016/08/03/progress-in-the-shift-to-conversational-computing>

¹⁰ <http://venturebeat.com/2016/06/14/what-300-million-users-taught-kik-about-chatbots>

via written language, but with advances in speech recognition, that increasingly seems a rather spurious differentiator.

By this definition, chatbot applications have been around for a long time. Indeed, one of the earliest NLP applications, Joseph Weizenbaum's Eliza, was really a chatbot. Eliza, developed in the early 1960s, used simple pattern matching and a template-based response mechanism to emulate the conversational style of a non-directional psychotherapist.¹¹

The fact that people seemed to be fooled into thinking that Eliza was a person rather than a machine inspired a whole community of interest in building chatbots that might one day pass the Turing Test.¹² This area of activity found popular expression via the somewhat controversial Loebner Prize,¹³ which, since 1991, has taken the form of an annual contest designed to implement the Turing Test. There has been a general tendency in academic circles to look upon the Loebner Prize with some disdain; Marvin Minsky once called it 'obnoxious and stupid'.¹⁴ But there's no doubt that the current hot commercial opportunity owes something to that legacy. It's worth noting that A.L.I.C.E. (aka Alicebot), a three-time Loebner winner, was built using the Pandorabots API mentioned above.¹⁵

There are other highlights, or perhaps lowlights, in the history of chatbots that we won't dwell on here—like the much-loathed Clippy the Office Assistant, which shipped with Microsoft Office from 1997 to 2003.¹⁶ Some might not want to call Clippy a chatbot, since it didn't really converse in natural language, but the basic UI paradigm is remarkably similar to that used in some of today's chatbot toolkits, with system responses often taking the form of stylized menus of options for the user to select from.¹⁷

But, from where I sit, the most relevant piece of retro technology that has resurfaced in the chatbot world is the finite state dialog modeling framework used in the speech recognition industry, made popular by VoiceXML in a series of standards since version 1.0 in 2000.¹⁸ Interacting with some chatbots is incredibly reminiscent of telephony-based spoken language dialog systems from the early years of the millennium, right down to the 'Sorry I didn't understand that' responses to user inputs that are out-of-grammar, and the sense that you're being managed through a tightly controlled dialog flow via the demand that you select your response to a system question from a narrowly proscribed set of options. I wouldn't be surprised to find that some of the chatbots out there are being built using a text interface to a VoiceXML interpreter, a functionality that's often available for testing dialog

¹¹ <http://dl.acm.org/citation.cfm?id=365168>

¹² <http://m.mind.oxfordjournals.org/content/LIX/236/433.full.pdf>

¹³ <http://loebner.net/Prizef/loebner-prize.html>

¹⁴ Quoted in http://www.salon.com/2003/02/26/loebner_part_one.

¹⁵ <http://alice.pandorabots.com>

¹⁶ <http://knowyourmeme.com/memes/clippy>

¹⁷ Also, I wanted an excuse to link to this video: <https://www.youtube.com/watch?v=LUvcYP859fg>.

¹⁸ <http://www.voicexml.org>

designs. Absolutely nothing wrong with that, and it's not a bad way to leverage tried and tested technology.

4 Why now?

So if there's not much really new here in terms of the basic technology that's being used, why the sudden commercial interest?

Well, a major factor is that the world has changed, and in particular, the way that people communicate has changed. When Weizenbaum introduced Eliza in 1966, interactive computing via a teletype keyboard was a new thing. Fifty years later, 6.1 billion people, out of a total human population of 7.3 billion, use an SMS-capable mobile phone.¹⁹ Messaging apps are used by more than 2.1 billion people: forty-nine per cent of 18–29-year-olds, thirty-seven per cent of 30–49-year-olds and twenty-four per cent of those over 50. Facebook Messenger alone has 1 billion users. We are entirely comfortable communicating via short typed interactions, and quite unfazed by carrying on several asynchronous conversations at the same time.

So, the big change here is the availability of a massively popular platform that appears to be an almost perfect environment for chatbots. Because of its ubiquity, the messaging interface is effectively a frictionless interface, just the kind of thing the Zero UI movement has been banging on about for a while now.²⁰ No more need to download, install and open up an app just to order a pizza; your conversation with the pizza bot has more in common with the texting you participate in to decide where to meet your friends after work. It's just another facet of today's always-connected multi-tasking world, where we participate in multiple conversations in parallel, each one at a pace of our choosing. Very soon we'll be in a world where some of those conversational partners we'll know to be humans, some we'll know to be bots, and probably some we won't know either way, and may not even care.

5 Where next?

There's often something of a disconnect between the questions we ask in our research labs, and the questions that need to be answered in order to build viable products. Of course, it's entirely right and proper that research should be ahead of the curve, and there are all sorts of reasons why we should pursue research whose immediate commercial benefit is not clear. But there are consequences to not paying attention to the potential connectivities between the longer term concerns of academia and the more immediate needs of industry.

In the language technology business, a prominent instance of a disconnect I'm aware of relates to the finite state dialog modeling work mentioned earlier. In the early 2000s, this was simply the only way you could build a spoken language

¹⁹ The statistics in this paragraph all come from <http://www.openmarket.com/wp-content/uploads/2015/10/OpenMarket-Portio-Research-SMS-the-language-of-6-billion-people.pdf>; other surveys will likely offer different numbers.

²⁰ <http://www.fastcodesign.com/3048139/what-is-zero-ui-and-why-is-it-crucial-to-the-future-of-design>

dialog system if you wanted it to have any chance of working with real users, and it's probably still the main paradigm used in that world, augmented with more statistically driven call routing algorithms. But while the commercial speech recognition world was developing VoiceXML, research labs were building much more sophisticated dialog systems that did much more sophisticated things, like unpacking complex anaphoric references, or trying to reason about user intentions. These systems worked great in carefully scripted demos, but you wouldn't want to mortgage your house to build a start-up around those ideas. As far as I can tell, not much of that work found its way into practical applications. In the world of spoken language dialog systems, there seemed to be two quite separate universes of activity, with relatively little awareness of each other.

Fast forward to 2016, and I see the risk of something similar happening in the chatbot world. In their striving to move the technology forward, the next milestone the Big Four are set to tackle is around truly conversational interactions, by which is meant the ability to take account of discourse context, rather than just treating a dialog as a sequence of independent conversational pairs. So you'll see this snippet repeated endlessly on blogs that discuss the capabilities of the new Google Assistant:²¹

These advances basically add context to your questions. For instance, when you say 'OK Google' followed by 'What's playing tonight?', Google Assistant will show films at your local cinema. But if you add 'We're planning on bringing the kids', Google Assistant will know to serve up showtimes for kid-friendly films. You could then say 'Let's see Jungle Book', and the assistant will purchase tickets for you.

Of course, the computational linguistics community has been looking at discourse phenomena like these more or less since the inception of the field, so it would be nice to think that the capabilities we'll see tomorrow on our phones will be informed by some of that research. That's already likely in the case of Google and the other major players, of course, since they hire people with that kind of background. I'm less confident that the broader chatbot-building community will have easy access to the relevant expertise. Although I agreed with Minsky's dismissal of the Loebner Prize at the time, I now think it may have unnecessarily alienated that community, and so some bridge-building might be in order, lest we end up with another pair of parallel universes.

And there are encouraging signs. In September, IVA2016 will host the Second Workshop on Chatbots and Conversational Agent Technologies.²² There's something conciliatory about using both the more formal—'conversational agents'—and the informal—'chatbots'—monikers in the same breath. Alongside the usual discussion of how we should best prepare for future capabilities, we might hope that the challenges of building a real fieldable chatbot will also get an airing there. Coming from the other direction, the World Wide Web Consortium has recently announced

²¹ For example, at <http://www.pocket-lint.com/news/137665-what-is-google-home-how-does-it-work-and-when-can-you-buy-it>.

²² <http://www.aclweb.org/portal/content/second-workshop-chatbots-and-conversational-agent-technologies>

a new W3C Community Group on Voice Interaction, which intends to explore beyond the system-initiated directed dialogs of the VoiceXML world.²³ This looks like a timely recognition that the earlier approaches taken in commercial systems have limitations that need to be transcended, and an excellent opportunity to revisit how some of the ideas developed in earlier dialog systems research might influence practical developments.

If we want to have better conversations with machines, we stand to benefit from having better conversations among ourselves.

²³ <https://www.w3.org/community/voiceinteraction>