7 August 2019

AI-Voice for everyone.

Introducing the Open Voice Network.

Artificial intelligence-enabled voice assistance is one of today's most transformative personal technologies.

It will soon become a primary means through which the world interacts with the internet. It will be resident on every digital device, from smart speakers to smart phones, from automobiles to appliances. In its ease, convenience of use, and eventual ubiquity, it will re-shape human communication and consumption behaviors.

It is also a technology that brings with it important questions of access and commercial freedom, personal privacy, and data ownership and control.

Similar questions have emerged in the past in periods of breakthrough technology innovation. In such times (such as at early days of the internet), the development of globally-adopted standards led to rapid growth and widely-distributed value creation. **We are in such a time with AI-voice. au**

For this, and additional important reasons, we announce The Open Voice Network. A non-profit association of enterprises, developers, and researchers, brought together in the belief that the future of AI-voice **must be like the internet, and thus *open*: standards-based, accessible to all, interoperable, and data-protected.**

Contents **page**

1
With the "zero interface," you just talk.

Artificial intelligence-enabled voice, simply stated, is the marriage of natural language recognition and computationally-demanding, data-hungry artificial intelligence. It is realized today in the chatbots that answer frequently asked call center questions, and in the voice assistants that work inside smart speakers , smartphones, automobile dashboards, and every-day home and kitchen electronics, from televisions to coffee makers.

AI-voice provides what analysts describe as the "zero interface" to the internet and to Internet of Things (IOT) systems. AI-voice users no longer type on a keyboard, nor swipe or tap on a touch-screen. To connect with each other, listen to music, explore the internet, or guide the functions of an automobile or appliance, ***they simply talk*.**

And, when required, the devices talk back. This is enabled by ever-more accurate recognition of human voices (across multiple languages and accents), the use of artificial intelligence to determine meaning, and powerful computing that can find accurate answers in a flash across the realm of the internet. And, by the ability of computers to create human-sounding responses.

For users of AI-voice assistants, this portends a behavioral revolution. At a command, the internet or an IOT system is explored or directed in an easily consumable form – be it news, music, a shopping list to be updated or confirmed, airline flight schedules, or the affirmation of a garage door closed. No longer must the internet be explored through a corridor of passwords, app downloads, Google page scans, and URLs. No longer must one wait on hold to access the information kept captive in call centers. Or, as an employee, pick an order in a store while holding a sheet of paper.

AI-voice makes it all here and now. In terms of consumer ease and convenience, it represents not a step, but a leap ahead. With AI-voice, the internet (and all connected devices) now stands at the consumer's beckon call: ever-present, eager to please, with the voice of a gracious, caring woman, and the manner of an ever-smiling concierge or English butler.

Or, potentially, the particular voice of a trusted brand.

2
The AI-voice revolution is in its earliest of days.

The technologies that enable AI-voice assistance were first explored in 1964 at the Massachusetts Institute of Technology. However, it was Apple's introduction of its Siri AI-voice assistant platform in 2011 and Amazon's introduction of the Alexa AI-voice assistant platform and Echo smart speakers in 2014 that has pushed AI-voice to the digital center stage.

Measured by consumer adoption, the smart speaker has been the **fastest-ramping device in consumer electronics history.** Roughly 75 million smart speakers were shipped worldwide in 2018, and a five-year (2017-2021) forecast of unit shipments suggests a global consumer adoption CAGR of 32 percent. In the United States (the world's leading AI-voice assistant market), NPR and Edison Research reported that smart speaker device ownership grew 78 percent in 2018 – from 66.7 million units to 118.5 million units. (Nearly one in ten US residents received a smart speaker as a 2018 Holiday gift.) Looking ahead, smart speakers are expected to be in at least 55 percent of US homes in less than five years.

However, the smart speaker phenomenon is but the front edge of **a much larger AI-voice assistant wave**.

- AI-voice assistants can reside on any digital device – in an era where nearly every device is digital. The leading device for AI-voice assistance is currently not the smart speaker, but the everyday smartphone. As of Q3 2018 in the United States, more than 90 million adults (35.7 percent of the adult population) were monthly users of AI-voice assistants on smartphones, a number nearly twice that of monthly users of AI-voice on smart speakers. As of mid-January 2019, more than 1 billion devices worldwide were equipped with Google's AI-voice Assistant. The UK's Jupiter Research has forecast that the number of voice assistants in use this year (2019) worldwide will total 3.25 billion– a number which will rise to 8 billion by 2023.

- The 2019 CES Show – the world's leading exhibition of the digital future – was filled with AI-voice assistant interfaces in everything from connected homes to connected cars. As of year-end 2018, nearly twenty firms around the world – including many leading e-commerce, technology, and telecommunications giants – offered or planned to offer proprietary, non-interoperable AI-voice voice assistant platforms and affiliated devices.

  Such firms are now working to embed their platforms into four primary ecosystems: personal mobile devices, home-based devices, vehicles, and workplaces. In time, we can expect AI-voice assistants to guide smart homes, automotive comfort-entertainment-direction finding, and workplace conference calls and unified communication – as well as our interactions with the internet, providers of goods and services, and each other.

Two other factors make clear that AI-voice assistance is in its relative infancy.

- **The functionality of today's AI-voice assistance**. Today's AI-voice assistant experience might be best described as one of "robotic Q&A." Orders are given or questions asked; the AI-voice assistant responds. As we look forward, guided by current research, it's clear that "robotic Q&A" is the first of many phases of functionality.

  The next phases – unfolding now -- will introduce unique, brand-specific voices that might be cheeky, chatty, or *basso profundo* wise, as well as the first steps into human-to-machine *dialogue*. This, in turn, will lead to more playful *conversation* that will enable, for example, more effective search by voice. In time – as suggested by what Google has shown in their recent "Duplex" voice assistant demonstrations -- we will begin to see the emergence of a fully

personalized, empathetic, anticipate-your-needs personal assistant, one that searches, schedules, and shops on your behalf.

This will not happen easily. Those versed in the mechanics of artificial intelligence know that moving AI-voice into dialogue and conversation will require veritable mountains of voice data – and armies of technologists to guide learning and inference. It was reported in November 2018 that Amazon had some 10,000 employees working on Alexa – and that may be a fraction of the AI-smart workforce required to drive AI-voice functionality forward.

- **The emergence of voice-vision functionality**. As noted, today's AI-voice functionality is best suited to specific commands and questions. Open-ended search through voice-only assistants – for instance, "*what are the best hotel options in Boston two weeks from tomorrow?"* – can be difficult, if not impossible. However, voice-vision combinations, as realized through the addition of AI-voice assistance to smartphone applications, or through a smart speaker-screen device, are expected to ease and speed search via voice.

3
AI-voice will re-shape consumer-brand relationships, shopping, and enterprise productivity in consumer-facing industries.

The use of AI-voice for the acquisition of information has pushed the technology into the world of consumer-facing industries and commerce.

Broadly stated, there are now three foundational value propositions for AI-voice in consumer-facing enterprises.

- Over the long term, the most important may be the one now beginning to emerge: **the use of AI-voice to create ever-growing intimacy between a brand and a consumer**.

  The prescient 2017 "Speak Easy" report from J. Walter Thompson's Innovation Group spoke to likelihood of voice users developing strong emotional connections to their AI-voice assistants. JWT researchers found that the emotional attachment to an AI-voice assistant grew over time, and predicted that AI-voice's transition from robotic Q&A to actual dialogue would rapidly deepen such bonds. According to the JWT report, 43 percent of regular voice technology users were said to love their AI-voice assistant – even to the point of wishing it to be a real person.

  For brands, this suggests an opportunity to create an ever-closer, trusted relationship with a consumer. Marketers have long spoken of a brand's *tone* or *voice*, which imparts the brand's positioning and personality. In print, brand voice is expressed not only through word choice, but in the selection of fonts, images, and design elements. Taken together, these communicate – oft intangibly – that a brand is authoritative, youthful, mom-centered, or cheeky.

  AI-voice assistance takes brand voice to a next level. Not only is there an actual voice – male or female? Authoritative or cute? – but AI-voice assistance is *interactive*. The brand now has the

opportunity to *listen,* and to millions, if not billions of users -- a sampling well beyond that of focus groups and surveys both in number and in immediacy. For brands, AI-voice listening can be unfiltered and in the moment.

Because it can listen, the brand also has the opportunity to *respond* – in an increasingly personal way. Here lies the value of applied artificial intelligence, as learning and inference can work at group, cohort, and individual levels.

And, the brand can now become a source of trusted and broadly-defined value. Here, we must envision AI-voice beyond its current robotic Q&A reality. The JWT report said it clearly: "brands that develop useful content can meaningfully engage with their consumers." Might the brand be a source for product usage tips and hacks? A source for parenting-centric podcasts aimed at young mothers? A place of entertainment and shared stories, all relevant to the brand promise? The place through which is hired a digital-virtual assistant?

in recent months, Convrg, the Los Angeles-based AI-voice and messaging company, developed for The GRAMMYs an interactive voice-based trivia game, music news tidbits, and artist sound-bites. For a brand such as Estee Lauder, they created a voice skill that delivered product recommendations and guided meditation.

- A second value proposition – also enormously significant -- is the **use of AI-voice assistance by consumers in the selection, purchase, and post-purchase management of products and services.** In short, *AI-voice for shopping* – whether for groceries, mortgages, or pediatric care.

  Since the emergence of AI-voice technologies, retail and consumer goods industry seers have predicted a new era of so-called "conversational commerce," wherein shoppers would explore and purchase products via AI-voice assistance. The data suggests that we are in the early stages of this new era. Although actual *purchases* by voice in 2018 were minimal, voice-centric *shopping* – i.e., interaction across the decision journey with a source of desired products or services -- began this past year to emerge in importance.

  - According to a January 2019 primary study of US voice assistant adoption and usage by Voicebot.ai, 28 percent of **smart speaker owners** -- a cohort of roughly 18.6 million digital, upper income individuals -- used their voice assistant regularly in 2018 to **search for products and services**.
  - The same Voicebot.ai study found that 26 percent of **smart speaker owners** -- a cohort of roughly 17.2 US million shoppers -- used their voice assistant regularly in 2018 to **find recipes or ask for cooking instructions;** 15 percent of smart speaker owners – a cohort of some 10 million shoppers – used their voice assistant regularly in 2018 to **purchase products or services**; and more than 30 percent of smart speaker owners wanted to use their voice assistants to **get answers from customer service departments**.
  - A November 2018 Voicebot.ai study of voice assistant usage on smartphones found that 28.8 percent of respondents used a smartphone-based assistant on a monthly basis to find a place to eat. The same study found that roughly one-quarter (24.8 percent) of respondents

had used a smartphone-based assistant on a monthly basis to research a product before purchase.

- The May issue of **Harvard Business Review** (*How Voice Assistants Could Change the Way We Shop*, Kane Simms, HBR, 15 May 2019, https://hbr.org/2019/05/how-voice-assistants-could-change-the-way-we-shop) noted not only AI-voice's convenient appeal ("everything is just an utterance away"), but the above-the-norm power of voice to convert leads to sales

- A third value proposition – and one of immediate value to the consumer-facing P&L – is the use of AI-voice for enterprise **operational productivity.** This is now being realized through:

  - **Voice-controlled employee and task management systems**. The U.S. Department of Labor estimated in 2017 that 60.4 percent of all work across US civilian employment was spent standing or walking – not sitting at a desk.

    In consumer-facing industries – and especially in retail and health care – the number may be higher. Many store- and patient-facing tasks demand access to digital information – and, the ability to use both hands to complete the task. AI-enabled, voice-controlled employee systems enable both.

  - **Chatbots.** These are automated, artificial intelligence-based programs – now most often used in call center or customer service situations – that respond to frequently asked questions with a simulated human voice.

    Unlike human employees, chatbots can work 24-7, and are resiliently accurate. A number of documented cases show that chatbots, when properly trained (in terms of artificial intelligence) and with access to necessary data, can become the preferred service source for human customers of retailers and consumer goods manufacturers.

    Like AI-voice assistants, the functionality of chatbots will grow in sophistication over time. Unlike AI-voice assistants, however, chatbots are accessible through open, interoperable communications systems, such as the internet or land-line or cellular telephone service.

A number of major consumer-facing firms, ranging from Bank of America (Erica, its AI-voice chatbot, with six million users as of March 2019) to McDonald's (testing voice assistants in drive-through services as of June 2019) to Procter & Gamble (Alexa skill for Tide Stain Remover and Oral-B dental care), Estee Lauder, Unilever, Target, and Walmart are now developing, testing, and implementing consumer-centric voice assistants.

This is a trend that is expected to accelerate, and not just because of the availability of AI-voice technologies and assistants. With the advent of e-commerce and the smartphone, there has been an accelerating worldwide consumer adoption of commerce-shopping concepts that offer greater ease, simplicity, and utility – in a word, **convenience.** This shift to convenience is reflected in the continued growth of e-commerce (estimated for 2019 by eMarketer to increase worldwide in revenue by 20.7 percent) and will be accelerated by societal transformations such as rapid urbanization, evolving gender

roles, redefined and smaller households, and generational shifts (especially that from aging baby-boomers to digitally-native millennials and Gen Z's.)

The shift toward convenience has significant implications for all consumer-facing industries. In retail, it will drive demand for online ordering and home delivery, leading to new fulfillment capabilities in physical stores. Within the fast-moving consumer goods (FMCG) segment of the industry, weekly "stock-up" trips to the grocer will gradually be replaced by online ordering or voice-driven "home replenishment" ordering – the latter enabled by a simple command of "*ship it again*."

For consumer goods companies, the equity of a brand may be challenged by the one-click ordering ease of a competitive, private-label product. Equally, the ease through which a brand provides valuable knowledge and delightful engagement (as well as great products) may supersede the temptations of a proprietary platform.

Simply stated: the consumer demand for convenience will not abate. It will seek, find, and use the easiest means to do the most.

And first among those will be AI-voice.

4

The emergence of AI-voice assistance as a primary interface to the internet – and, as a multi-function platform -- raises a host of important issues.

As Voicebot.ai's Bret Kinsella recently noted, AI-voice assistants serve as both interfaces to the internet and IOT systems, and as platforms for a growing list of services.

As interfaces to the internet and IOT, they enable (if the right words are said in the right sequence) navigation, task execution, and information retrieval. As such, they are the *entry valves* of the voice-enabled internet – turning on or off at the request of the user or according to the self-interests of the provider. Wrong words? Off. Competitive product or platform? Off.

As platforms, they create ancillary business services, and encourage the development of third-party applications – and do so on an ever-growing foundation of user voice data. Their relative market strength is reflective of, and determined by, the number of supported services, size of third-party development ecosystem, and weight of data for artificial intelligence learning and inference.

At present, a handful of companies now dominate the voice assistant and smart speaker markets with *proprietary* interfaces and platforms. In general, these interfaces and platforms are not interoperable. They can block consumers from competitive brands and products, and those brands from consumers. In addition, as platforms, they claim and gather data from most interactions, including those of third parties.

And, as AI-voice grows in availability and adoption, the proprietary interfaces and platforms are positioned to extract an ever-greater value from both consumers and all consumer-facing industries.

Given the small number and large size of today's proprietary AI-voice interface and platform gatekeepers – and the fact that (as noted by the JWT report), for a vast majority of users, *voice-only internet search works best when it provides just one answer*. This suggests that we are moving beyond search engine optimization to a new, expensive era of pay-to-play recommendations and *algorithm optimization.*

There is another issue of importance – one that touches the commercial and personal realms of AI-voice usage. Voice offers a remarkably rich vein of data to analysts and data scientists. As such, it *expands the range of personally identifiable information (PII).*

As we know, voice communication is far more than spoken words. Voices also express meaning through tone, pitch, loudness, and rate. It's why pauses, verbal stumbles, and changing intonation make a voice assistant sound "human" – and why robotic responses sound artificial.

The characteristics of tone, pitch, loudness and rate can be used to identify individual voices, as well as indicate emotional states (such as anxiety, calm, confidence) and even physical and mental illness.

Today's AI-voice environment lacks standards, reference implementations, objectivity, and associated services. The experiences that a user can create and provide through the current AI-voice ecosystem is limited by and unique to the AI assistant being used.

It is a situation that recalls the early days of the internet.

The internet (and its public face, the World Wide Web) has evolved from a series of protocols that enabled the connection of independent government and academic computer networks into today's ubiquitous, impactful presence in nearly every facet of developed world life.

Its path from lab project to a primary tool for billions of humans was driven *not* by a central governing body, nor by the efforts of a dominant technology company. Instead, the internet was birthed and grown through the development (and adoption) of **technology and governance standards** that enabled ***interoperability*** and ***domain ownership.*** Those, in turn, sparked availability, adoption, and unceasing innovation.

For instance: the Transmission Control Protocol/Internet Protocol (TCP/IP) developed over the years 1973-1978 made it possible to join almost any network together. In 1982, the "IPv4" 32-bit numeric web address standard was set forth by Internet Engineering Task Force (IETF), which made possible nearly 4.3 billion discrete web addresses. At roughly the same time, it was clear that centralized management of numeric addresses would be impossible to manage – which led to the development of the **Domain Name System (DNS),** with original specifications published by the IETF in 1983.

The DNS is a hierarchical and decentralized naming system for computers, servers, and other resources connected to the internet. It translates readily memorized "domain names" to numerical IP addresses, and as such, serves as the "telephone book" for the world wide web. Registries of domain names are

accredited by the Internet Corporation for Assigned Names and Numbers (ICANN), or other similar standards-based organizations.

Given the importance of technology and governance standards to the growth (and value) of the internet and World Wide Web, **what of AI-voice?**

To foster the *broadest-possible adoption, innovation, and value-creation*:

> **Might we need a standards-based address and wake-word registry for AI-voice – one that is akin to the Domain Name System (DNS)?**
>
> **Might we need standards that would enable and encourage interface and platform interoperability?**
>
> **Should access to the marketplace of goods, services, and ideas be gated by the interests of proprietary interface and platform providers?**
>
> **Should the data of that marketplace – data that ranges from personally identifiable information to commercial data of products, pricing, and promotion -- be claimed (and owned) by proprietary platform providers?**
>
> **Might we need new approaches to the acquisition and use of personally identifiable voice data?**

5

Leading academic researchers are now exploring a future of AI-voice that is *open*.

In 2016 researchers at two prestigious U.S. universities – the Massachusetts Institute of Technology (MIT), and Stanford University – began, independently, to study AI-voice. And, specifically, the questions listed above.

The MIT effort is centered in the Auto-ID Laboratory, known for its leadership in the development of the now-global Electronic Product Code (EPC) standard for RFID-based product identification in commerce industries, and its ongoing advisement of Global Standards One (GS1), the worldwide organization governing product identification and commerce process collaboration. The AI-voice work noted here is led by Professor Sanjay E. Sarma and Professor Brian Subirana, and is aimed at developing a standards-based approach for AI-voice that is "open" – defined as interoperable, accessible, and protective of personally-identifiable and commercial data.

The Stanford research has its home in the university's computer science department and is led by Professor Monica S. Lam. In May of this year Dr. Lam and team received a $3 million grant from the National Science Foundation for the pursuit of their vision for Almond, an "open, crowdsourced, privacy-preserving and programmable virtual assistance for online services and the Internet of Things."

The MIT research began with a three-way exploration (with the Intel Corporation and Capgemini) as to the opportunities and implications of AI-voice in commerce, which led to the January 2018 publication of a white paper ("Time to Talk".) Later that year, the Target Corporation provided seed funding to an Auto-ID Center team to begin the study of a standards-based approach to AI-voice, one that might lead to future of interoperability, data protection, and accessibility with and through any AI-voice interface and platform.

At the center of the MIT Auto-ID Laboratory research is the initial design of a potential **Voice Name System (VNS) and VNS wake word construct and registry.**

The Voice Name System – similar in intent to the Domain Name System (DNS) of the internet – would enable the reservation and use of "wake words" to activate AI-voice interfaces and platforms. A "wake word" is a one or two-word verbalized phrase that turns on a voice assistant platform; in the envisioned VNS, it would work much as a URL works on the internet: a door to a clear, unfiltered connection to a site of choice. Such a system would standardize voice interactions to a universal reach (similar to systems such as telephone numbering). In addition, the envisioned VNS would be endowed with "wake neutrality," which would allow each participant could securely customize the voice, personality, and "wake word" of their AI-voice experience.

The Voice Name System is also envisioned to offer a set of standard commands, definitions, settings, and modes – as well as standards for data protection.

The initial research also suggests that a central and standardized repository of wake words be established and managed by a neutral third party, such as a university or global standards body. It could offer trained machine learning models that enable speech-to-text processing, as well as a cross-reference against a repository of previously-registered wake words.

Given the ambiguity of voice and the fact that speech recognition is not 100 percent accurate, techniques used in previous standards (such as domain queries for the DNS of the internet) would fail to adequately capture the nuances of voice and overcome challenges such as pronunciation, accents, language and context. As such, a VNS standard may need to define several ways of dealing with error correction, including requiring access by spelling, confirmation prompts, or verification on a different device. It may also need to assert geographic limits (in the US, it's Kohl's – in Australia, it's Coles), or require other elements added to a wake word ("OK Coles") to assure identification, as well as define boundaries for similarity and overlap of wake words.

The MIT research toward a Voice Name System is also addressing issues of data security and privacy. Current research has shown that wake algorithms can be fooled by sound sequences inaudible to the human ear, or be modified to operate as wiretaps – opening the door to hackers and limiting trust. The use of voice as biometric authentication aggravates the dangers of insecure listening devices. The risk of identity theft also increases.

The MIT research team has also developed a sample architecture which could grant AI-voice enterprise users integrated access to Global Standards One (GS1) product reference data such as the Global Trade Item Number (GTIN) and associated product attributes.

At Stanford, the questions noted above have led Dr. Lam and team to the development of a neutral AI-voice interface and assistance platform – one designed to be open, intelligent, social, and private, and to enhance open competition. This platform, named "Almond," was introduced in 2017. To further the work, Dr. Lam this year formed at Stanford the Open Virtual Assistant Lab (OVAL), consisting of colleagues in computer science, medicine and law, as well as off-campus technology leaders.

According to Dr. Lam, the mission of OVAL is to advance virtual assistance technology, with special attention to its human impact. In a recent presentation, she noted both the compelling value proposition of virtual assistants (enabling the use of voice to access web services and Internet of Things data) and their power (accessing data, and serving as an intermediary between a consumer and digital services. As such, *proprietary* AI-voice interfaces and platforms are a potential threat to consumer privacy and open competition.

To protect privacy – and enable social connection – the Almond architecture does not centralize data in one site. Assistants are run locally on devices, and share information with each other without a third party viewing or scraping data. This is especially important for the protection of personally-identifying and commercial data in commerce, health and life sciences, and financial services. It also offers the prospect of liberating more data for AI-based research across all entities.

The architecture of Almond is comprised of three core components: *Thingpedia*, a public knowledge base of natural language interfaces and open API's to services and devices which enables Almond to connect broadly; *ThingTalk*, a high-level language that can connect web services and Internet of Things (IOT) devices in a single line of code; and, *ThingSystem*, an open-source system that manages user credentials and data, and executes *ThingTalk* programs.

The *ThingTalk* language was invented by Dr. Lam and team. It has a simple construct; users can say "when" something, "get" something, and/or "do" something in a single message, and have the specific actions parsed and prioritized appropriately. When connected to the endless reach of *Thingpedia*, users can easily connect to services and, just like they do in today's type-tap internet, aggregate the knowledge they want.

Dr. Lam's commitment to openness is not limited to the Almond platform. The actual artificial intelligence, the neural model, has been made completely public, so that enterprises can embed the natural language interface in their own products and voice services. *Thingpedia*, LUInet and all the platform tools are also publicly available. In addition, the Stanford team is addressing the critical issue of data acquisition and AI training in an open, non-proprietary way. The mastering of human language, with its complexity and variability, will demand massive data sets. Thus, Dr. Lam and team are pursuing "open" data training development – enlisting not tens of thousands, but hundreds of thousands, even millions in the work of creating correct training samples.

6
Why "open" AI-voice will be adopted by consumers.

Given the market leadership in North America and Western Europe of the leading big-tech proprietary platforms, it is reasonable to ask if and why AI-voice consumers might choose to use or fully adopt an alternative, "open" platform or a standards-based approach.

Skeptics can point to not only the market share, but the "stickiness" of legacy AI-voice platforms. As AI-voice use increases across the breadth of services now offered across some highly-integrated proprietary platforms, one can expect that habit loops will form, convenience will rule the day, and the cost of change will be too high to overcome.

However, there are several reasons why consumers will eventually lean toward – and then adopt – an open, standards-based approach.

- **Interoperability.** In an AI-voice world that connects consumers to the internet and IOT services, interoperability will matter. Yes, there will be consumers who will be content to live their AI-voice lives within the bounds of a proprietary platform. However, we are facing an AI-voice world not of 100 million smart speakers, but of some *eight billion AI-voice assistants by 2023*. This is a world in which AI-voice enabled ecosystems include automobiles, entertainment, home security, home appliances, health care, public safety – and shopping. This is a big world – one owned not by one or two platforms, but by up to twenty global platforms. In this world, there is a clear need for platform interoperability.
- **An open, standards-based approach to consumer technology is good for digital consumers.** Mozilla's Firefox browser, for instance, delivers both a usage alternative and a policy- and market-shaping influence to the commercial market leaders.
- **The robust assemblage of AI-voice services and applications that inevitably emerges in "open" ecosystems**. No one or two companies – massive as they may be – will be able to (or find it valuable) to create a veritable universe of AI-voice applications. As demonstrated with the internet (and mobile internet), an "open" development community, working on standards-based platforms, can and will for AI-voice.
- **A demand for high quality and consistency of voice recognition**. AI learning and inference demands data. AI learning and inference for the complexity of voice demands *mountains* of data. The transition from "robotic" command-and-response to contextually-aware and nuanced dialogue will demand the acquisition and assessment of enormous data sets. In such a situation, an open, crowd-sourced approach may be better positioned for success than a closed, proprietary effort.
- **Widely-dispersed brand trust.** Consumers trust different brands, and for many different reasons. That trust – and the retail-brand products, expertise, and consistent provision of satisfaction and delight – will not go away. Consumers will want direct and unfiltered access to those brands.

- **Standards can and will be adopted by existing proprietary platforms.**
- **Privacy will matter on private matters.** Multiple recent studies have shown that today's digital users – despite observed behaviors to the contrary -- show significant concern for data privacy. As usage of AI-voice moves in coming years beyond early adopters to the early majority, the issue of personal data privacy will raise its head.

  It is reasonable for today's consumers to worry little about the privacy implications of music requests to Pandora or Spotify, sports trivia questions, requests for jokes, or queries about the morning weather or commute. However, it will be a much different thing -- as the number of users increases, usage increases, and AI-voice functionality increases -- to ask the internet questions about the health of a child or sexual health. To check bank account or credit balances. To order prescriptions or merchandise that, for whatever reason, is desired to be known only to the consumer and the supplier.

7
Forging the way to open AI-voice: The Open Voice Network

In January 2019, representatives from leading retailers, consumer goods companies, industry associations and marketing firms met in New York to discuss the potential, opportunities, and challenges of AI-voice in consumer-facing industries. From that initial gathering has emerged The **Open Voice Network** (OVN), a not-for-profit association **dedicated to AI-voice that is open: standards-based, interoperable, accessible, and data-protected.**

In May 2019, the Open Voice Network was invited to become a **Directed Fund program of The Linux Foundation.** The affiliation of the Open Voice Network with The Linux Foundation reflects a mutual commitment to open, standards-based technology development, and will afford the Open Voice Network access to the shared services of The Linux Foundation while enabling organizational strategic and decision-making independence.

The Linux Foundation has membership of more than 1000 enterprises, and is dedicated to building sustainable ecosystems around open source projects to accelerate technology development and commercial adoption. At present, The Linux Foundation is the organizational home of such projects as Open Network Automation Platform (ONAP), Hyperledger, Cloud Native Computing Foundation, Cloud Foundry Foundation, Xen Project, and many others.

Support for the work of the Open Voice Network is being provided by consumer-facing enterprises that share the concerns – and the mission – of the OVN. As a Directed Fund program of The Linux Foundation, the OVN will minimize operational costs through the use of Linux Foundation shared services, yet maintain decision-making and directional independence.

Given its mission of open voice – standards-based, interoperable, accessible, and data-protected – the Open Voice Network will create open AI-voice value in three ways:

- **Standards development:** through third-party research and recommendations toward the global standards that will enable interoperable, accessible and data-protected AI-voice assistant search, communication, and commerce.
- **Industry value and awareness**: identification and sharing of voice-centric value propositions and best practices across consumer-facing industries; serving as the source of insight and interpretation for AI-voice in consumer-facing industries.
- **Advocacy:** working with and through existing industry associations on relevant regulatory and legislative issues, including those of data privacy and antitrust.

In addition to supporting the above, OVN member firms will also be invited to participate in industry symposia that will speak to OVN-sponsored research and best practices in AI-voice commerce. Research topics are expected to include not only standards development, but studies of voice-in-commerce consumer behavior and future use cases.

Given the pressing issues of voice search and commerce, initial membership is focused upon the retail and consumer goods industries. It is expected that membership will be expanded to other consumer-facing industries within 2020-21, such as the media, health and life sciences, automotive, telecommunications, and smart & connected cities (including public safety.)

For more information:
Jon Stine, Open Voice Network
+1 503 449 4628
jon.stine@openvoicenetwork.org
www.openvoicenetwork.org

Open Voice Network
23 July 2019
8.0

Selected Bibliography

A level playing field for the Digital Single Market, Ecommerce Europe, 2019

Anand, *The Reality Behind Voice Shopping Hype,* The Information, 6 August 2018.

M. Ceglowski, *Statement of Maciej Ceglowski, Founder, Pinboard, Before the U.S. Senate Committee on Banking, Housing, and Urban Development On the Topic of "Privacy Rights and Data Collection in the Digital Economy,* May 2019.

Chatterjee, *Give Your Brand A Voice – Literally,* Forrester, 23 July 2018.

Consumer Intelligence Series: Prepare for the voice revolution, PWC, March-April 2018.

Consumers will increasingly use their voice assistants to shop . . . , Business Insider Intelligence, 10 July 2018

Conversational Commerce: Why Consumers are Embracing Voice Assistants in Their Lives, Capgemini Digital Transformation Institute, January 2018.

Conversational Commerce: The Rise of Voice Assistants, eMarketer, May 2018.

Creswell, *Amazon, the Brand Buster*, New York Times, 23 June 2018.

Dawar and Bendle, *Marketing in the Age of Alexa*, Harvard Business Review, May-June 2018.

Domain Name System, Wikipedia, 2019.

Engelson, *The Future of Voice: From Smartphones to Smart Speakers to Smart Homes*, comScore, 15 December 2017.

Engelson, *Smart Speaker Penetration Hits 20% of U.S. WiFi Households*, comScore, April 2018.

History of the Internet, Wikipedia, 2019.

Khan, *Amazon's Antitrust Paradox,* The Yale Law Journal, 2017.

Kim and Duhachek, *Artificial Intelligence and Persuasion: A Construal-Level Account,* 2019.

Kendall, *Conversational Commerce in China*, Lewis, 12 September 2018.

Kinsella, *Amazon Echo Device Sales Break New Records,* Voicebot.ai.com, 26 December 2018.

Kinsella, *84 Percent of Businesses Expect to Use Voice Technology with Customers in the Next Year,* Voicebot.ai.com, 17 June 2018.

Kinsella, *RBC Analyst Says 52 Million Google Home Devices Sold to Date and Generating $3.4 Billion in 2018 Revenue,* Voicebot Weekly, 24 December 2018.

Kinsella, *Google Assistant Now Available in Brazilian Portuguese on Android, iOS and Even Google Home,* Voicebot.ai.com, 19 November 2018.

Kinsella, *Think Tank – Are Duopolies the Natural Order of Voice?,* Voice Insider, voicebot.ai, 20 November 2018.

Kinsella, *Think Tank – One Voice Assistant or Many?,* Voice Insider, voicebot.ai, 8 November 2018.

Kinsella, *Voice Insider #18 – Why Smart Speakers in Italy and Spain Were More Important Than Google Duplex in 2018,* Voicebot.ai, 28 December 2018.

Kinsella, *Think Tank – Privacy as a Voice Assistant Feature,* voicebot.ai, June 2019.

Kinsella, *Think Tank – Is Voice an Interface or a Platform? Yes, of course it is,* voicebot.ai, July 2019.

Kinsella, *Think Tank – A trillion-dollar economy? A voice assistant peace dividend? Where is all this headed?,* voicebot.,ai, July 2019.

Lam, *Presentation notes,* 2019.

Lam, Campagna, Xu, Fischer, and Moradshahi, *Protecting Privacy and Open Competition with Almond: An Open-Voice Virtual Assistant,* 2018.

Lam, Campagna, Ramesh, Xu, and Fischer, *Almond: The Architecture of an Open, Crowdsourced Privacy-Preserving, Programmable Virtual Assistant,* International World Wide Web Conference Committee, 2017.

Listen, and learn, The Economist, 27 October 2018.

Maheshwari, *Marketing Through Smart Speakers? Brands Don't Need to Be Asked Twice," The* New York Times, 2 December 2018.

Markoff, *Stanford Team Aims at Alexa and Siri With a Privacy-Minded Alternative,* The New York Times, June 2019.

Marchick, *Voice Search Trends*, Alpine.AI, 15 April 2018.

Meeker, *Internet Trends 2018*, Kleiner Perkins, May 2018.

Meeker, *Internet Trends 2019,* Bond, July 2019.

Mehta, *Indian Voice Assistant Developed by IIT Patna Will Speak 18 Languages,"* beebom.com, 11 December 2018.

Metz, *Finally, a Machine That Can Finish Your Sentence,* The New York Times, 18 November 2018.

Mockapetris and Dunlap, *Development of the Domain Name System,* Computer Communications Review, Vol. 18, No. 4, August 1988.

Natural language understanding, citing *Brief History of AI,* American Association for Artificial Intelligence," Wikipedia, updated June 2018.

Petrock et al., *Dr., Alexa Will Hear You Now,* eMarketer, December 2018.

Petrock et al., *What's Next for Voice Control? Digital Assistants, Smart Speakers and the IOT*, eMarketer, June 2018

Policy recommendations on the role of online platforms in the e-commerce sector, Ecommerce Europe, April 2016.

Physically strenuous jobs in 2017, The Economic Daily, U.S. Department of Labor, Bureau of Labor Statistics, 2018.

Rowe, *Building Trust in an Age of Zero Interface*, Publicis Media, 27 June 2018.

Schumpeter: Firefox and Friends, The Economist, 20 July 2019.

Simms, *How Voice Assistants Could Change the Way We Shop*, Harvard Business Review, 15 May 2019.

Smart Speaker Consumer Adoption Report, voicebot.ai and Voicify, March 2019.

Smart speakers are the fastest-growing consumer tech; shipments to surpass 50 million in 2018, Canalys, 4 January 2018.

Speakeasy: the Future Answers to You, Mindshare Futures, J. Walter Thompson Innovation Group, 2017.

Speech and Voice Recognition Market Worth $31.82 Billion by 2025," Grand View Research, November 2018.

Subirana, Puig, Gaidis, Tarrago, Soens, Rajesekaran, Stine, Cantwell, and Sarma, *RFC: The MIT Voice Name System (VNS) White Paper (draft),* June 2019.

Subirana et al., *AI-VOICE Platform Capabilities Development Forecast, 2017-2025,* Auto-ID Laboratory, Massachusetts Institute of Technology, December 2017.

Switched-on Consumers – Their Data, Their Terms, Data IQ Research Report, 2018.

Taylor, Jacobs, Warner, Graman, Sarma, Subirana, Cantwell, Hunt and Stine, *Time to Talk,* January 2019.

The Digital Assistant of Tomorrow, Juniper Research, 2019.

The Quest for Convenience, The Nielsen Company, 2018.

The Smart Audio Report, NPR and Edison Research, June 2019.

The Smart Audio Report, NPR and Edison Research, January 2018.

The Smart Audio Report, NPR and Edison Research, Spring 2018.

Voice Assistant Consumer Adoption Report, voicebot.ai, Pullstring, RAIN, November 2018.

Voice-Enabled Commerce: Insight Report, Gartner L2, October 2018.

23 July 2019
8.0