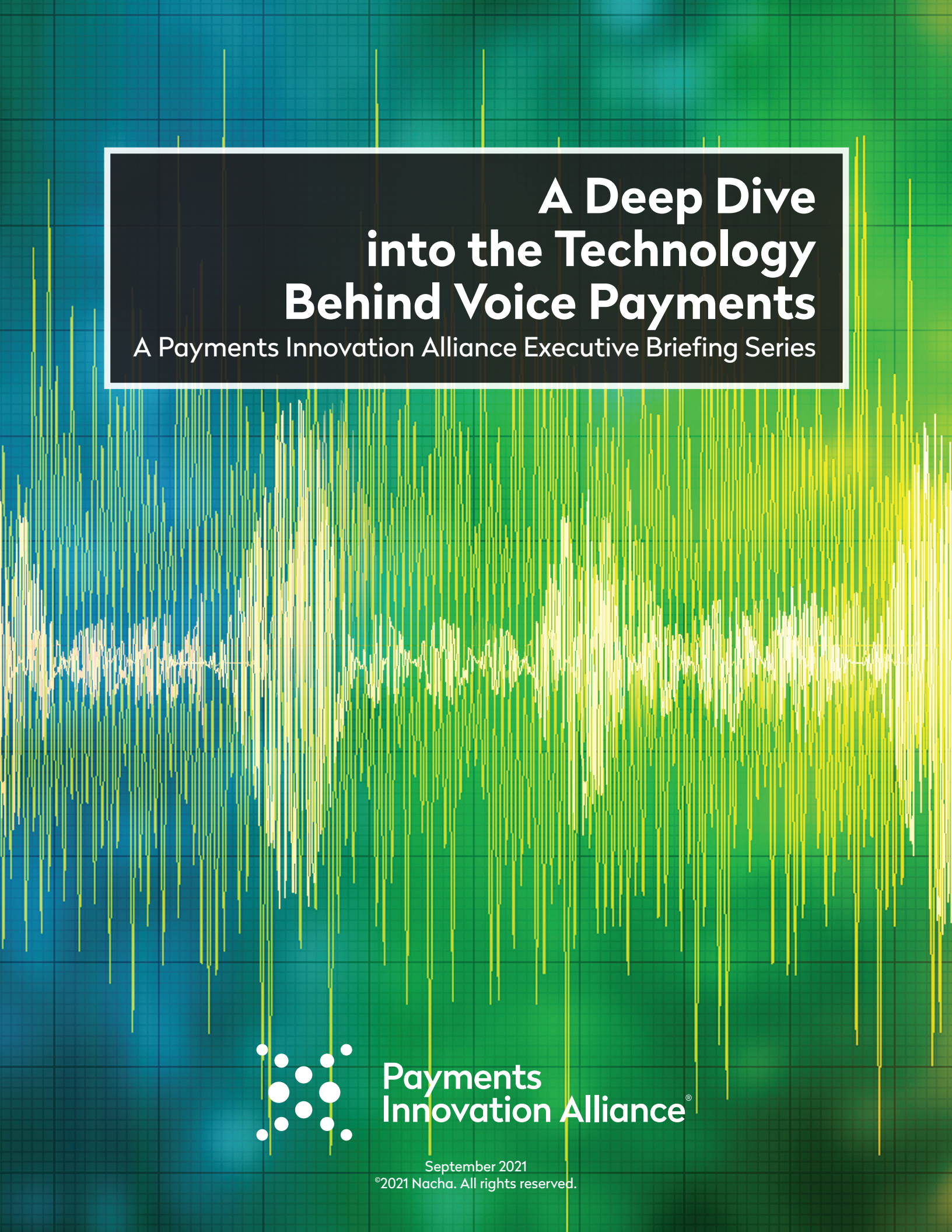
A Deep Dive into the Technology Behind Voice Payments

A Payments Innovation Alliance Executive Briefing Series



Payments Innovation Alliance®

September 2021

Introduction to the Executive Briefing

This Executive Briefing Series consist of a set of articles published by the Payments Innovation Alliance. Future articles will cover a wide range of topics relevant to financial services' participants and will be published on an ongoing basis. For more information about the Payments Innovation Alliance or the Conversational Payments Project Team, visit www.nacha.org/payments-innovation-alliance.

This article is the second in the series.

History

Today's voice payment technology stems from the advancements made in artificial intelligence (AI) and machine learning. AI refers to the capabilities of machines being able to problem solve and respond in a similar manner as a human would. Machine learning is the process by which a machine develops its AI.

Learning algorithm programs are created and used in modeling processes in which scientists teach AI how to respond correctly, and adjust when it goes off course. The objective is for the machine to respond back in an expected way. Dating back to the 1950s, scientists such as Alan Turing, John McCarthy, and others first started exploring the mathematical capability of building intelligent machines and developing the concepts and framework for what was to come next in AI.

In 1955, the first AI program "Logic Theorist" was created by Herbert Simon, Allen Newell and John Shaw. This was the first program of its kind developed to conduct artificial reasoning.[i] Later on, programming languages (such as LISP) were developed, advancements in learning algorithms were further enhanced, and technology such as artificial neural networks were created allowing machines to further mimic the process of thinking like a human brain does. These developments are the backbone of AI today.

In addition to AI, voice assistants (such as Amazon's Alexa or Google Home) utilize the power of voice recognition technology. Dating back to the 1950s, voice (or speech) recognition technology was developed to translate spoken words into numbers that computers could understand. Today, voice assistants utilize voice recognition technology to help them better understand the commands of the specific person that the assistant is listening to. They record and catalogue that information and then create profiles that help the assistant more easily recognize the person's voice for increased accuracy in future queries.[ii]

* Key terms used in this briefing are defined at the end of the document.

Components

Put plainly, a voice payment is the process of speaking to an AI unit, such as a smart speaker or smartphone, with a request to make a payment or to buy a product or service. Utilizing AI technology, it can recognize the verbal request, act on the request, and respond back confirming what it did. The four primary components of AI that allow this to work are machine learning, natural language processing, robotics, and computer vision.[iii] Each of these components work together to accomplish the ability for machines to respond and think.

Machine learning

- Modeling used to train AI to learn how to problem solve
  - Supervised learning - human gives AI the correct answer and trains it to respond that way
  - Unsupervised learning - Trial and error process designed to help AI arrive at the correct answer on its own
- Both processes are not perfect and require multiple iterations

Natural language

- AI learns to understand what humans are communicating and how to respond
- Accomplished through training processes much like how a child would learn to read and write
- Natural language processing is the key component to how voice payments are made possible

Robotics

- Mimic the physical form and can automate tasks and even problem solve
- Used in industries such as manufacturing, logistics, health care, and even in banking in varying ways

Computer vision

- Process in which AI learns to recognize images
  - May be accomplished through massive amounts of images teaching AI to recognize objects such as flowers, animals, cars, etc.

Current Capabilities

Voice payments are made possible by using artificial intelligence, specifically tapping into natural language processing, which allows AI to communicate with us. This technological combination is referred to as conversational AI, which opens the possibility for chatbot technology. Chatbot technology, which is the premise behind digital assistants, is used to leverage AI in solving problems through communication.

Some call centers and websites use this technology to automate customer service requests. In those instances, the chatbot will use natural language processing to understand what the person is requesting, tap into its learning algorithms to determine how to respond to the request, and again use natural language processing to provide a response. This conversational AI is the mechanism behind voice payments. Your request to make a payment, purchase through an online seller, and more goes through a similar process.

Digital assistants such as Google Assistant, Alexa, and Apple's Siri have the potential to facilitate a payment providing they are connected to the parties involved in the transaction, meaning, the payment can be done via voice if parties of the payments participate and/or allow the payments to be made in that manner.

Are voice payments differentiated from other types of payments such as online, phone, or in person? The answer is no. Voice payment uses whatever is already connected to the merchant (for instance, your Amazon account) to make the payment. In that way, it looks and acts more like an online or mobile payment. These transactions are not directly controlled by the financial institution. Instead, the digital assistant leverages the internet to gain access to the sites as if it were the user doing it directly.

The Internet of Things (IoT) allows for many possibilities for voice payments to be used in many more places. No longer are we tied to desktop computers when we need access to the internet. Devices such as smartphones, thermostats, security systems, gym equipment, and even refrigerators are connecting to the internet, too. As we see these smart technologies further integrate into our homes, offices, cities and even cars, the idea behind a mobile payment really starts to take a new form.

Assistant	Activation Speech	Examples of Compatible Devices
Alexa by Amazon	"Alexa"	Echo, Echo Dot, Fire TV, Fire Tablet, Ecobee Smart Thermostat, Ring Smart Doorbell
Google Assistant	"Ok, Google" "Hey, Google"	Google Nest Hub, Google Nest Mini, Nest Thermostats, Android Smartphones, Pixel Slate tablet, August Smart Lock
Siri by Apple	"Hey, Siri"	Apple HomePod, MacBooks, iPads, iPhones, August Smart Lock

User Experience

Voice payments can be made through financial institutions, billers or through device providers. The user experience varies depending on where the payments are being initiated.

Payments through device providers. Providers such as Apple, Amazon and Google each offer solutions for consumers – creating multiple smart speaker ecosystems. The biggest challenge with multiple providers is that each system operates independently of one another, which means that users need to either choose one brand and stick with that, or they will have multiple setups with devices that may not be able to connect with one another because the interfaces and standards for each system may be different. All these factors create a less ideal user experience and cause confusion and frustration. Similar to disparate wallets, smart devices are not interoperable. For example, if a user's preferred mobile and computer platform is Apple, and they have a Google Home device, access to third-party commerce sites (e.g., Amazon) would need to be set up separately because the devices and environments are not currently interoperable.

As with the various payment rails, different providers mean the operating rules for these systems may also differ with user and merchant participation from network to network. There is also the question of encryption, tokenization, authentication, and access controls, which again may vary with each ecosystem.

Payments through financial institutions or billers. Direct biller payments are still leading financial institutions in overall bill payments. This is due to the fact the end consumers do not receive the type of visibility from their financial institutions as they would from their direct billers site (such as bill details, card and ACH payment methods when payments post to the account). As a result, billers want to create and control the user experience. On the other hand, financial institutions are banking on the convenience of making multiple payments to drive engagement on their site. Through a direct biller, a user can only pay one bill at one time, while through a financial institution website, one can add multiple billers into the site and make payments as invoices are received.

Today, there are a growing number of financial institutions and billers offering voice payments as part of their bill payment services. Given the complexities and security challenges described above, some may support only one or a couple of digital assistants. The benefit with these services is twofold. First, it is meeting the market need of a service already being used by consumers for other types of payments. Second, unlike the limited controls in place for other payment types, financial institutions that enable voice payments through bill pay services have the ability to control how payments operate by requiring users to create the initial bill payment setup using a mobile device, tablet or computer, as well as requiring the user to authenticate on the smart device – including requiring multifactor authentication – as part of the bill payment process.

In summary, it is important to remember that transactions conducted through voice-enabled devices are already occurring both outside and inside of the direct control of financial institutions. While access to proprietary systems (such as bill pay and online banking) can be controlled by financial institutions, access to the Amazon Marketplace, to third-party biller sites, and for other web-initiated purchases are tied to the internet. If voice-enabled devices are able to connect them (like Alexa can connect to the Amazon Marketplace), users are already able to conduct e-commerce using them. [iv]

U.S. Smart Speaker Consumer Adoption Report for 2019 revealed that 15% of U.S. smart speaker owners say they were making purchases by voice on a monthly basis at the end of 2018.[v] That is up from 13.6% that were using voice for retail purchases at the beginning of the year. This reflects a 10.5% rise in the relative use of smart speaker-based voice purchasing on top of a 40% rise in year-over-year device ownership. It is anticipated that consumer expectations to conduct financial transactions such as obtaining account balances, making transfers and bill payments using smart speakers, will grow as a natural extension to e-commerce.

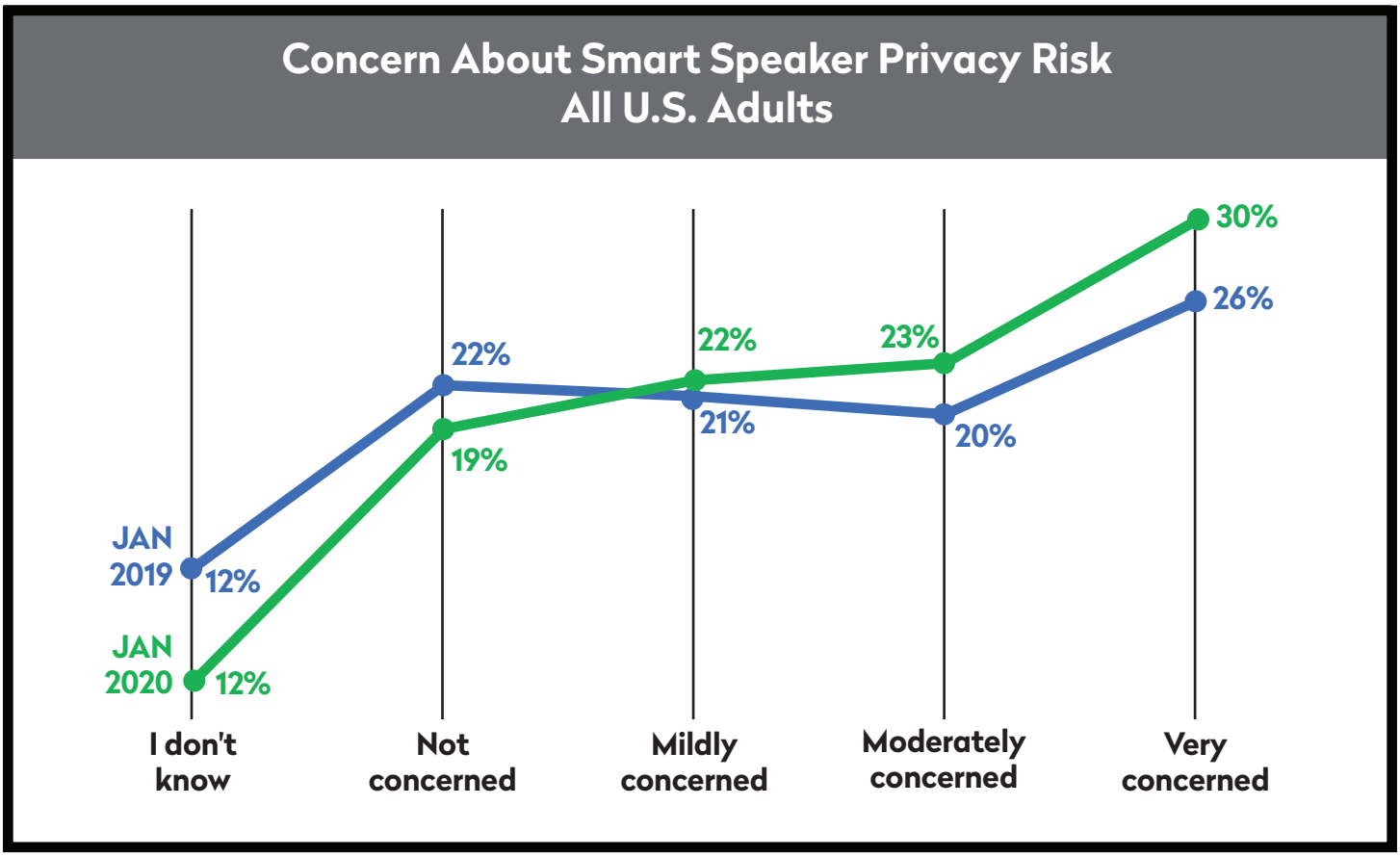
Limitations

Privacy and security. Even though voice assistants communicate with their servers using encrypted connections, there is still a concern about privacy and security in general, and voice payments as an emerging use case.

Listening to conversations to identify errors and make improvements in machine learning models for voice assistants is a common and necessary practice. Most device providers store the audio recordings to personalize a user's experience or to serve them advertisements. Without this, the systems do not improve speech recognition or identify user intent as quickly. This is a risk for voice assistant providers because errors in natural language processing have a direct negative effect on user experience.

In the last couple of years, media coverage exposed that conversations were being recorded and consumer privacy concerns increased.[vi] Not only did the coverage position the listening as something that was kept secret by voice assistant providers, but it revealed that third-party contractors were doing the work and potentially leaking private information to consumers.

Advocates have called for more transparency about how these companies use customer data. It is important to note that even for bill payments made using a smart speaker through a financial institution or a biller website, the device may be storing the audio recordings.



Source: Voicebot.ai 2020

Additionally, the potential exists for a transaction to be initiated by someone other than the authorized party. Using smart speakers could increase a user's vulnerability to voice hack, a subset of identity theft in which someone obtains an audio recording of the user's voice and uses it to access their information. Once they have this recording, they use it to trick authentication systems into thinking they are the user. This hack is a potential way to get around smart speakers' voice recognition capabilities.

Smart home speakers provide a potential goldmine of audio recordings that someone could use for voice hacking. If a bad actor manages to hack into the speaker or cloud service where a user's records are stored, they could use it to hack into various accounts.

While some smart speakers provide controls to delete collected data and recordings, without additional controls (such as biometric authentication, multifactor authentication, secondary controls, etc.) it is entirely possible that anyone could use this technology to perpetrate fraud providing they have access to the device. This is an area that does not have clear standards yet.

AI Considerations. AI uses learning algorithms (computer programs) to train it to respond in the correct manner. Unfortunately, this process is not infallible. Put into the wrong hands, AI could be trained to respond in an incorrect or malicious manner. There is the potential that, if programmed accordingly, the voice assistant could be used to access a user's information to steal money or perpetrate fraud. This lends itself to the ethical considerations behind AI technology in its totality. In the scope of voice payments, there needs to be a degree of inherent trust in the technology itself. If a person is at home and wants to quickly purchase something on Amazon, he or she should be able to trust that doing so will not compromise his or her accounts. This is where the controls and understanding of how the voice payment technology is derived becomes a crucial element in its ongoing usage.

AI is only as good as the data scientist programming it to learn or act. The potential for AI to be taught to discriminate (i.e. declining all loan applicants based on sex or race) is very much a possibility. If the criteria it uses to make loan decisions lines up with characteristics of a specific race, age, sex, etc. it could violate lending laws. This could be a side effect of AI learning to act. In addition, since AI is a tool, it can be used to perpetrate fraud. Imagine an AI programmed by fraudsters to steal a few dollars disguised as a fee from each P2P transaction it sends out. These are just two examples, but serve to illustrate the potential concerns around the use of AI.

The Future

Looking to the future, a pressing question is what kind of advancements are on the horizon that will further expand the capabilities of voice payments?

Privacy and security: All smart technology comes with privacy and security risks. That does not necessarily mean these devices should not be used, but when it comes to financial transactions, keeping customer data safe is vital. Financial institutions and billers should take steps to implement privacy and security controls, and help educate users on the measures they can take to protect themselves. Today, consumers are initiating both e-commerce and voice payments using smart devices without strict security controls in place. While there is a concern that adding friction to the payment can reduce the benefits gained from the medium, such as voice payments, consideration does need to be given as to how these payments can be secured.

It can be expected that while additional controls may be put in place, there will be enhancements to the technology itself to be more foolproof. One example is the voice assistant responding only to the voice of its owner utilizing voice recognition profiles. Fraud detection and prevention services, such as positive pay type services, are another option. The industry will drive how to address security concerns, including sensitive data storage, as voice payments technology evolves.

Smart device standards. There is currently no industry standard for smart devices. However, in December 2019, Amazon, Apple, Google and the Zigbee Alliance announced that they are working together to create a new, royalty-free connectivity standard to increase compatibility among smart home products, with security as a fundamental design tenet. The Zigbee Alliance project is built around a shared belief that smart home devices should be secure, reliable and seamless to use.[vii] By building upon Internet Protocol (IP), the project aims to enable communication across smart home devices, mobile apps and cloud services, as well as define a specific set of IP-based networking technologies for device certification. The goal is to decrease friction in the current environment and provide for a better user experience.

5G Impact. 5G, which stands for fifth generation of cellular networks, may provide the infrastructure to make internet driven technologies faster and increase their capacity. 5G speeds are up to 10 times faster than those of 4G. This means that more connections can be made, and the access and response time of 5G downloads will be much faster. This opens the possibility of improving AI in that its success is dependent on how quickly it can process and respond. With 5G, that response time may become much faster, which opens the possibility of additional technological advances, including fully functional smart cars that can respond to situations in a similar manner to a human-controlled vehicle.[viii]

This increase in capacity and processing ability brings us to the intersection of AI, IoT, and 5G and paves the way for smart cities. While some smart cities exist today, a faster infrastructure can facilitate the growth of more smart cities and existing smart cities can expand their capabilities. These cities utilize IoT, AI and other technology to automate functions and improve the overall quality of life for those residing in them. The idea is that technology can help streamline the experiences of people within cities such as public transit (using technology such as voice payments and near-field communication to conduct frictionless transactions), securing neighborhoods and automating security scans (AI and robotics usage), and even sustainability through technology use in growing food (AI and robotics usage). [ix]

While the results of which are not fully known, the trends we are seeing regarding technology capabilities and the potential enhancements of the future, lead to the speculative conclusion that the future of voice payments may become reality as smart technologies, faster processing speeds, and technology standards allow for continued integration into the varying aspects of our lives – despite privacy and security concerns. While we do not know how this will impact the way we transact payments in the future, we know for sure that our options are expanding.

Key Terms

Artificial Intelligence (AI) – The concept of an intelligent (thinking) machine

Artificial Neural Network – A group of computers (nodes) that work together in a connected manner that mimics the structure of a human brain

Biometric Authentication – Using human anatomy (such as fingerprints) to gain access to a device or system.

Chatbots – A form of AI that can take text or audio commands and respond to and act upon thema

Computer Vision – Ability for machines to read and understand images

Conversational AI – A discipline of multiple technologies (including voice recognition, chatbots and software) to personalize communication between a machine and a human.

Digital Assistant – This is a computerized program that provides answers to questions, access to databases (such as music or web searches), and can take action (such as by conducting a requested transaction) for the user (e.g., Siri or Alexa)

Internet of Things (IoT) – The concept of varying types of devices (such as smart speakers, smartphones, thermostats, alarm systems, etc.) connecting to the internet

Learning Algorithms – Program used in machine learning to teach a machine to problem solve

LISP – Programming language (developed for AI use)

Machine Learning – The process by which a machine is trained to learn and improve from problem solving

Multifactor Authentication (MFA) – Utilizing two or more authentication methods to gain access to a program or device

Natural Language Processing – The mechanism that allows for AI to learn how to understand and respond to the human language

Near-field Communication – The ability and process for two devices to connect to each other when in close proximity (e.g., Apple Pay)

Robotics – Use of machines in physical form (e.g., assembly line automation machines, or HSBC's banking assistant robot)

Smart Cities – The integration of cities with technologies (such as IoT) to automate and streamline processes and experience

Smart Speaker – A smart speaker contains a built-in microphone that listens for wake words, such as "Alexa" or "Hey, Google." Voice recognition programing uses hot words to execute commands such as performing tasks, making payments, setting appointments and reminders, or linking to websites. Smart speakers may also be referred to as voice activated speakers or voice speakers.

Smartphone – A mobile phone that performs functionality typically found on a laptop or PC. There may be a touchscreen interface, voice activation, smart speaker, internet access, and ability to run downloaded applications.

Supervised Learning – The process in machine learning in which machines are trained to pick the correct choice from predetermined solutions

Unsupervised Learning – The process in machine learning in which machines learn to recognize patterns and come up with solutions on their own

Voice Payments – Technology that allows a user to request a transaction verbally. This is typically done in conjunction with a digital assistant-enabled device (e.g., smart speaker)

Voice Recognition – The ability for a computer or machine to understand a voice request. Pre-programmed commands are spoken and the software will execute them.

About Nacha

Nacha governs the thriving ACH Network, the payment system that drives safe, smart, and fast Direct Deposits and Direct Payments with the capability to reach all U.S. bank and credit union accounts. Nearly 27 billion ACH payments were made in 2020, valued at close to $62 trillion. Through problem-solving and consensus-building among diverse payment industry stakeholders, Nacha advances innovation and interoperability in the payments system. Nacha develops rules and standards, provides industry solutions, and delivers education, accreditation, and advisory services. Learn more at Nacha.org.

The Payments Innovation Alliance is a diverse membership group of domestic and international organizations uniting to support payments innovation through discussion, debate, education, networking, and special projects. The Alliance incubates new ideas and initiatives to advance the industry. Visit https://www.nacha.org/payments-innovation-alliance.

For more information about joining the Alliance or the Voice Payments Project Team, please contact Jennifer West at JWest@nacha.org.

i Logic Theorist: Complete History of the Logic Theorist Program

ii https://history-computer.com/ModernComputer/Software/LogicTheorist.html

iii Link your voice to your devices with Voice Match https://support.google.com/assistant/answer/9071681?co=GENIE.Platform%3DAndroid&hl=en

iv Natural Language Processing (NLP): What it is and why it matters

v https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html#howitwork

vi Amazon Marketplace is an e-commerce platform owned and operated by Amazon that enables third-party sellers to sell new or used products on a fixed-price online marketplace alongside Amazon's regular offerings.

vii U.S. Smart Speaker Consumer Adoption Report 2019

viii https://voicebot.ai/smart-speaker-consumer-adoption-report-2019/

ix Amazon and Google are listening to your voice recordings. Here's what we know about that

x https://www.cnet.com/how-to/amazon-and-google-are-listening-to-your-voice-recordings-heres-what-we-know/

xi https://zigbeealliance.org/news_and_articles/connectedhomeIP/

xii https://www.cnn.com/interactive/2020/03/business/what-is-5g/index.html

xiii Smart Cities of the Future: From Vision to Reality https://www2.deloitte.com/us/en/pages/consulting/solutions/smart-cities-of-the-future.html?id=us:2ps:3gl:consem21:eng:cons:31020:nonem:na:6sKjOFEK:1177105429:424633433641:e:Smart_Cities:Vision_To_Reality_Exact:nb