ergomania

Voice is replacing point-and-click

Banking and Fintechs around the world are adopting voice

Consumers are increasingly comfortable with voice services

Voice creates opportunities to enhance existing offerings

Voice assists marginalized users

UX-UI has a vital role to play

The race for voice banking excellence

Transformation and change

In 2020 I was part of the Ergomania team writing a white paper titled ***Voice banking – transforming the financial experience***, exploring what was happening in the market. Revisiting the piece, I anticipated a few 'tweaks' here and there to update on the technology, and expand on the data.

Well yes, but no...

The time of the pandemic proved fertile ground for many developments in Fintech and banking, as working from home, and an increasing reliance on our various devices became ever more mainstream. Hurdles of trust and acceptance were rapidly overcome, and having being a niche and 'maybe soon' offering, voice use is now becoming more mainstream. So a quick rewrite of the existing whitepaper? Not quite.

Harlan Cockburn

Table of contents

Defining terms

Let's begin by clarifying what we mean when discussing how voice systems are utilized, as there are two distinct areas: **Voice recognition and voice activation**. With voice recognition the aim is to have a machine respond to fairly simple instructions such as opening a door, or playing a song. The application doesn't care about the gender or age of the speaker. As long as there is a reasonable understanding of the language, syntax, and relatively clear accent or dialect on the part of the user, then the system will usually deliver on the request. At the same time it will differentiate between a human voice and the vocalizations of other animals. In this way the family dog is screened from barking out requests to Alexa or Siri.

On the other hand, **voice activation dives deeper**. Our voices are as unique as our fingerprints, and now the goal is to individually recognize the person, rather than simply the words spoken. By initially reading a set text and recording it, **a voice recognition system - also known as voice biometrics - can be used as a unique personal identifier**. This has to work even if the speaker has 'flu, or is in a high ambient noise environment. When a voice recognition system compares a speaker with the samples collected in a database, it *should* be able to then accurately identify that individual. If it can match the speaker to the stored 'voiceprints' then it can be used as a tool for secure online banking.

We'll check in presently about how true this is.

However for now, touchscreens and a reliance on biometric fingerprint recognition are set to be superseded by voice use to identify customers, and give access to systems. Why? Because we all like the hands-free way of doing things, and how this allows us to simultaneously do other tasks**.** This isn't only about convenience however - it's more than just an alternative way of logging in, enabling a completely handsfree experience. Voice activation empowers and gives access to those who are visually impaired, or people who must work with full hands, such as medics needing to consult with colleagues while working. The pandemic has also pushed forward the demand for contactless, more hygienic interfaces.

In this white paper we'll be seeing examples of both voice activation and voice recognition. The boundaries are sometimes blurred, and applications are often hybrids of both, but the principles are essentially the same. Voice recognition and voice activation technologies use a combination of algorithms, artificial intelligence, natural language processing, and machine learning to recognize and respond to voice commands. Our interface with these systems is through smartphones, or the approximately eight billion 'voice assistants' now in use, such as Siri and Alexa. That's one device for everyone alive.

So how do we get our technologies to react to our voices, answer us, and then do what we want?

Understanding 'understanding'

András Rung is CEO of UX-UI Agency Ergomania, and with a PhD in Linguistics and Artificial Intelligence, has the skillset to address the subject of smart audio and voice-based systems. Specifically, Ergomania deals in interfaces for Fintechs and banks, and Andras is a frequent presenter at conferences where he helps explain the process of how we talk and listen to our devices. **The simple fact is, we are not consistent, and rarely frame our requests either concisely or logically**. For example, let's say I want to know when a cinema is showing a particular movie, or even if that movie is showing at all. I might ask my voice assistant: *When's Top Gun Maverick on?* or *What's on at the movies today?* or *Are there any seats available at the cinema?* All of them are valid questions, and if I was speaking to another human, they would quickly figure out answers, often by asking further questions. Or let's say I want 'A long black coffee' which is the same thing as an 'Americano', but with no common words. A human server in a coffee shop will quickly flex their understanding of vocabulary and syntax, but what does a machine do?

Andras Rung has some of the answers.

Who needs voice?

Starting from the start, Andras asks who needs voice applications, and the answer is that anyone can benefit from being handsfree and having quick and easy interactions with devices and systems. This can include being mobile in a car, or even while in the shower! There are many drop-downs of business values and advantages that come from voice use, **including the onboarding new customers and demographics, while improving engagement and even brand loyalty**. It can also attract Millenials and Gen Z-ers who are Digital Natives, and who expect nothing less than the most innovative solutions. It's less hassle and faster to talk than type, and even when using Messenger, some prefer to use voice instead of keying in text. So, do they want to deal with financial institutions still stuck in the previous century? No, they do not.

Then there's the cost advantages of using chatbots, designed to answer simple questions quickly and therefore saving on human staff costs. Not only is that much more efficient, but it means that when a customer really does need a human operator, they get to that point more quickly: **Less frustrating wait times, and better economies for the organization**.

OK, so there are clear needs and wants around voice use, but there are also a whole bunch of **challenges to be overcome**, not least in the way each of us speak. We have different accents, vocabularies, speed and articulation, and we don't always set the context of what we're talking about. The slowness of internet connection may sometimes bring proceedings to a near halt, and then there are issues of trust, and speaking in public places about sensitive issues such as our finances or health. But what's going on when we talk to machines (and they talk back)?

Language and tone

According to András Rung, a fundamental is to understand the language of users. As already mentioned, **we can ask for the same thing in different ways, using different vocabulary**. That means the voice assistant we are dealing with also has to 'understand' the different ways of being asked to do a task, and respond in a suitable way. An early part of any Ergomania project mapping out voice use is to define the personality that will be used, with gender, tone and other character traits defined. Will the voice be reassuring, warm or enthusiastic, cool or calm?

Although the processes are actually overlapping, let's view it linearly and say that next comes consideration of the Prompt Anatomy. This is the architecture of what can be presented to the voice system and in what order. Firstly, systems 'like' to have logical and sequential requests. Humans rarely provide this, at least to each other, and typically we 'talk around' subjects. For example, *Are you doing anything tonight?* is probably not a request for factual information, but more likely an invitation to meet up. Setting the context is therefore key to helping voice assistants get our message. *When does the next train leave for Vienna?* would be helped greatly by saying *When does the next train from Munich leave for Vienna?* **This leads to always decreasing the cognitive load on the system**. People have only a certain capacity to hold a given amount of information in their working memory at any one time. Bots aren't so different, so the prompt anatomy should be as simple and direct as possible. It's obvious when you think about it, and yet we expect the 'miracle' of AI to unravel all of our own tangled thoughts and instructions. It's therefore important we use well defined repair phrases that lead users back to a conversation and give the bot what it 'needs'.

People also do not all speak alike in terms of the patterns of stress and intonation they use, so the voice system must receive prompts across a wide variety of what is referred to as **Prosody**. That is how words are actually spoken aloud: *I want a cup of TEA* conveys a different shade of meaning to *I want a CUP of tea*. We use such stress patterns all the time in person-to-person communication, and the more successfully a voice assistant can respond to our prosody, the better. What's more, the bot should also use prosody to successfully communicate with us, **producing a more natural sound and experience**.

Much more than...

Let's move on to the **Cues** that form the communication between human and machine. Early systems used simple Yes/No responses which were themselves developed from employing alphanumeric keypads to request services. 'Do you want to check your balance?' requires an easy response of Yes or No and such menu driven approaches are common. Our physical location is also another typical cue to the bot to provide requested information, with the assumption that - for example - nearby restaurants are of more interest to us than those on the other side of the world.

Yes

&

No

And what specifically does UX-UI design bring to the party, with the underpinning theory of voice recognition and activation? Ergomania majors in research-led design processes which are arrived at through **integrated measurement, permeating all areas of a business, and mapping functions based on mental models**. In other words, products are developed not through endless arguments or ad hoc executive decisions, but based on the steady appliance of science. In terms of the development of Ergomania's **voice based banking research**, this begins with extensive fieldwork and interviews to establish the scope of a project, and a thorough grounding in Voice User Interface (VUI) technologies, plus Construction Grammar, a branch of **modern language theory**. This was pioneered by linguistics professor Adele Goldberg whose work has inspired analysis of grammatical constructions in different languages.

Charles Fillmore's work is another touchstone for the Ergomania approach, which incorporates his research into syntax and lexical semantics. For example, a single word, 'bank' can mean 'the edge of a river', or the action of an aircraft turning, or a place where money is stored. As we have seen, **context enables humans to understand how a word is being used at any particular time**. Machines have to be just as smart in order to support our wish to have safe, secure and easy to use voice based banking. (And note that now, 'bank' has turned from a noun into a verb: 'to bank').

Opening the 'Black Box' of Natural Language Processing (NLP) underpins many of the Ergomania processes of investigation and workshopping with clients. **NLP expresses the discovered working mechanism – the syntax – of what makes human speech work**.

Parsing

A typical analysis of a sentence could look like this: 'Alexa, ask Skill Master what are skills?'

Here *Alexa* is the 'wake word' which triggers a response. *Ask* is the 'launch', and *Skill Master* is the 'skill name'. *What are* is the 'utterance', and *Skills* is known as the 'Slot Value'. Having a frame construction with the phrase *What are…* (something), where in this case 'something' is skills, is a way of using construction grammar, where a value can be changed, based on 'rules'.

Having established *'What are…'* the rules allow a value such as *'skills'* (or 'dogs' for that matter), to be inserted. However *'because'* would be outside the construction grammar.

Similarly, for example, taking the Slot Value of 'the door' - based on both semantics and grammar, a reasonable assumption is that we can open, or close the door, but not *eat* the door. **Using a frame construction grammar allows a set of 'rules' to be applied so that within described boundaries only a specific number of actions can be applied**.

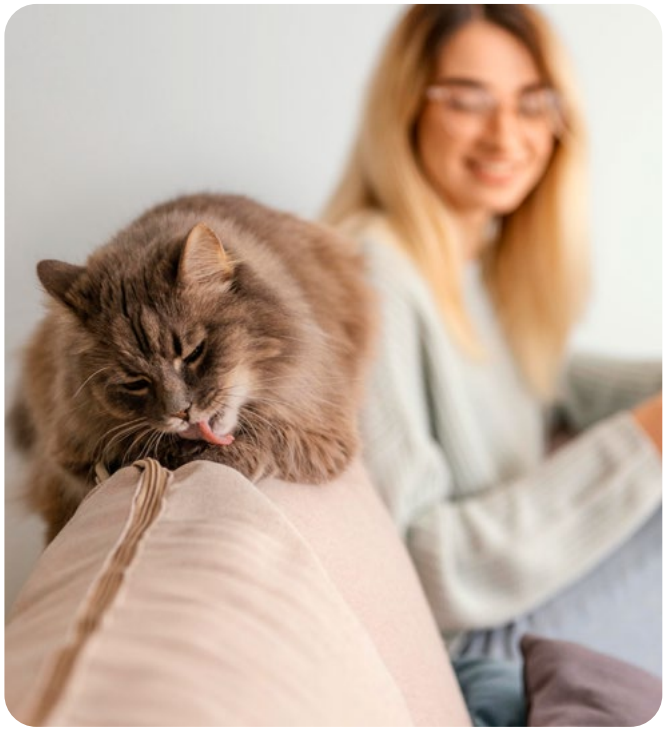
Or, to rephrase the interaction with Alexa, we could make the utterance, *'Tell me more about'* followed by the slot value of *'skills'.*

It's the kind of processing we all constantly do in any conversation, but **breaking down and understanding the syntax of a sentence requires a lot of work when designing a system to accomplish such transactions**. And although this may seem obvious, it's only through rigorous understanding and application of syntactical and lexical theories that Ergomania is able to bring added value to Fintech clients.

An early part of any Ergomania project mapping out voice use is to define the personality that will be used, with gender, tone and other character traits defined. Will the voice be reassuring, warm or enthusiastic, cool or calm? This also involves copywriting, where decisions must be made about the informal, or formal tone used, and what expressions are voiced. For example, does the bot say 'Hi', or 'Good day'? These factors also must be aligned with the purpose of the voicebot, and the branding of the company. Cultural factors play a part too, and the same bot should speak differently about the same topic to - say - Dutch, French or Japanese people. So what are the customers like who the bot will speak with… and um, what about cats?

Dialoging

Renáta Szilágyi-Nagy is a Senior UX designer at Ergomania, who began her explorations into the use of voice applications, and particularly dialog flow, by creating a *Cat Advisor app*. It was a fun test subject, using Google's voice platform to define what breed of cat best matches someone's needs.



Building on the experience gained with the cat project, Renáta and colleagues next tackled several internal projects before moving on to a money transfer Proof of Concept for a large European bank in Hungary - which won an innovation award for the internal team. Since then there has been a live voice project for KBC bank in Hungary, as a soft launch, rather than a 'big boom'. The KBC voice assistant, Kate, is also running in Czechia, Belgium, Slovakia and Bulgaria. Renáta does mention that the Hungarian population seems more receptive to voice offerings than other countries in the region. And from cats and money transfers using voice services, to the subject of Wildcards.

As we have seen, dialog between humans is not always 100% direct or clear, and **in practice we don't need to understand each and every word in a sentence to get the meaning**. If someone shouts, *There's been an accident, call a…* We will immediately deduce that calling for a Doctor or an ambulance is the most likely request, even if we miss that keyword. But how is this achieved within AI, when - it would seem - the system would either wait for further information, or would need to scan through endless permutations? In cases such as this, UX designers employ methodologies such as 'Wildcards'.

In coding, a Wildcard is a symbol such as an asterisk (*) or question mark (?) used to replace or represent one or more characters. The importance of utilizing wildcards in planning the workflow of a system is that a single character can be used to represent a whole string, when either it's not necessary to describe that whole string, or because it *can't* be described – there is too little information to identify the object or statement with 100% accuracy. So AI might be able to identify that it is being presented with 'animal' but might not be absolutely sure whether 'animal' is a dog or a cat. Time to deal the wildcard.

In 2022, in the USA alone around 42% of the population used voice at least once a month

And, *There's been an accident, call the ** would indicate an unknown, or uncertainty. We don't want a process to grind to a halt just because one factor is unknown, so the wildcard approach allows dialog to continue even without 100% 'understanding'.

Dialog Mapping is a further methodology used by Ergomania, where a flowchart is created to **graphically outline a process**, derived from a logical series of enquiries. If a user is asked the right questions, then they'll probably receive the appropriate answers. I might say that I want to travel to Stockholm. Your response could be 'Why?' My answer would be, 'To meet up with a friend'.
Dialog Mapping which resulted in the question, *'How* do you want to get to Stockholm?' would be immediately more useful, because I would be prompted to reply 'By air.' The next thing the system should be asking is, 'Do you want me to search for flights?'
Notice, by the way, that I used the **linguistically ambivalent term**, 'By air'. A very literal and rather OCD system would be trying to figure out how I would use 78% nitrogen, 21% oxygen and 1% other gasses to reach my destination. A *smart* system would know, or have learned that 'By air' can be interpreted as traveling in an airplane. It's key that systems must adapt to regular expressions. For example 'Call my bank' is typically used by American English speakers, while Australian and British English speakers may say 'Ring my bank'. The intent is the same in both cases, but the system must be responsive to such variations. Both terms must be in the frame. This applies to large-scale use, and down to personal interaction. If a typical opening gambit of mine is to say, *I'm wondering if...* then this needs to be recognized as a regular expression which is equivalent to the Alexa 'wake' word.

Who uses voice assistants, and what are the numbers? Well, in 2022, in the USA alone, over 142 million people used voice at least once in a month. That's around 42% of the population, but that's for all instances of voice use, with financial applications at the lower end of activity, being roughly 10% of the total. 2022 saw an overall modest rise of around 3.5 million users, and the projection for 2025 is 153 million U.S. users. This is still based on that 'once a month' assessment, although **it seems clear that as people become more and more comfortable with voice activation and voice recognition, the numbers will rise**. Even back in 2020 when Forrester surveyed the market for voice-enabled banking, the findings showed that 30% of adults with a bank account checked their accounts using a smart speaker, and the research suggested that, 'customers are starting to use them more for routine banking tasks.'

Theory into flows

This outline of the theory of voice systems hardly scratches the surface, but should be enough to establish the general principles, and the fact that this is a wide-ranging and complex subject. It's also tempting to refer to systems as 'understanding' us, which isn't really the case, but is a useful analogy to our normal human-to-human interactions. So to get a clearer idea of how to put theory into practice, I return to Renáta Szilágyi-Nagy at Ergomania, now a veteran of the KBC voice implementation in Hungary. She starts me off with the concept of 'Happy Conversation' which immediately sounds intriguing. It means our first drafting of a design conversation, which is 'happy' because nothing bad happens, Renáta explains.

So we're not talking about error handling here, and we're not at this stage thinking about what happens if we don't understand what the user says. The Happy Conversation assumes that everything will go right at the back end, with the user, and the language, and selecting the right use case. **Choosing the right use case is crucial as that will determine what the bot understands and how the user will be able to interact with it**. In this respect the current use of chatbots can be helpful, as insight is provided into what customers are asking for. Behind this is a team analyzing the kinds of questions being asked by real users. For instance, the most common questions are around the balance of an account, but there may also be upcoming new products, so how will these be introduced to customers, and what questions will those customers want to have answers for?

Knowing the market

This also ties into ongoing market research, and rigorously establishing whether proposed features will actually yield real benefits. Are real customers asking questions that need to be logged and understood? What are people interested in knowing?

I ask if this is the financial organization doing the research, or its UX partner? It's both says Renáta, and at this early stage of the *Happy Conversation,* **there will usually be a lot of workshops held to understand user expectations, and the technical perspectives of what the bot needs to do**, as well as the more 'esthetic' issues such as tone of voice and overall approach. As the workshopping process increasingly aligns client and customer needs, and defines the technologies to be used, sample dialogs and flow-mapping are started. The flowchart will have multiple - and complex - decision trees mapped out which are dependent on how a user may ask a question, and how many possible answers can be given.

For example, as a customer, I may ask to be told my bank balance, but the system 'knows' that actually I have *three* accounts. Do I want to know the overall total, or just a specific account? **The process of designing the flows is highly iterative, with the business and the developers constantly circling back to check and recheck until a good match is produced…** for now.

Putting Happy Conversations to the test

Testing now comes into play, to establish if the flows are working, and the question/answer decision trees are doing what they should. And this is all automated presumably? Renáta laughs - no, not really. English is more open to automated tools because of its dominance in markets, but one very useful approach is the *Wizard of Oz* technique, which requires no software, only very patient and precise people. **The wizard method is a sort of role play where one person acts as the customer, with another person being the bot, who can only respond with predefined answers**. The bumps in the road as customer and bot try to communicate are carefully logged by an observer to record which interactions work, and which don't.

As the Wizard of Oz process starts shaping up the 'right' kind of answers, the copywriting team gets onboard. Up to this point the project has been less about the *actual* wording, and more about the logic, but now copywriters begin to shape the dialog into more conversational forms. Did the initial briefing call for an 'uplifting', or a 'calm' tone? Clearly, users need to have a consistent experience with a voice bot which doesn't change from moment to moment in the way it talks, or shifts personalities.

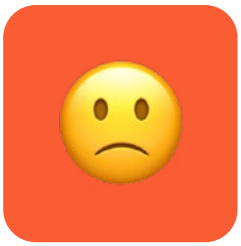
So if all that seems to be good, Artificial Intelligence training can begin, with massive datasets - the training phrase variations - fed into the system, constantly monitored by Natural Language Understanding (NLU) Developers. This allows the language models to be changed in parallel, so that the right sequences lead to the right intents. As a customer, my intent of being able to access the balance of my account may seem relatively simple, but we've already seen that may be made more complex by my having several accounts, credit cards, and so on. It's also unlikely that a bot will be created *solely* to check balances, so there will be many other functions which will be addressed by the NLU Developers, with repeated changes to the structure as ever more complex training words and sentences are introduced.

And we're still 'Happy' at this point?

Getting Unhappy

We're now approaching the 'long tail' of where the teams must start considering how to deal with flows which *don't* flow, for whatever reason. Are the questions not coming in right? Are the answers not being formulated properly? What happens if the required information can't be accessed from the backend? What if even, I as a customer am asking the wrong question to start with? Perhaps I have mixed up my debit account with my credit account - my bad, but can the system cope with that? With the *un*happy conversation flows launched, there are now many iterations of repair flows to be undertaken, covering everything from the language used, to backend development, and even something as simple as what happens when a connection goes down during a money transfer. It can happen and it *will* happen, so how can I as a user be reassured that all is well?

I'm describing this as a very *linear* process, but of course most activities are happening in parallel and in many layers, otherwise developing a voice project for banking or Fintech would take...



Well, how long?

How long is a piece of string?

The answer is, it depends. It depends on the releases within the bank - be they weekly, monthly, or on a longer cycle, and whether or not a bot is already up and running. For something as relatively simple as a Frequently Asked Questions application, it could be operational within a few weeks, because there is only one potential answer for every potential question. On the other hand, long use cases, such as account opening, may take many months to design because there are so many backend systems which need to be in place. There may also be custom UI elements to be designed, and constant checking, and more checking. **Are question inputs 'understandable' by the AI, and are the answers given understandable by customers? Is the language right, is the tone right?** Are all the 'unhappy' instances of conversations accounted for? And so on.

The answer is that the process is likely never ending, because even if a finite line could be drawn under one application, **new products and services will be introduced soon enough, which require further research, development and testing**. Even a use case which has already been launched may require further 'tweaking' as knowledge comes in of how it performs, or is being used.

The dominance of English

Using an approach such as the Wizard of Oz can be done in any language by the team conducting the development, but it has to be noted that throughout the world, English is the dominant business language with an inbuilt bias, as noted by Renáta Szilágyi-Nagy, whose initial Cat Advisor demo had to be in English, despite being tested with a Hungarian audience. This is of course because the giant tech companies are rooted in English, although it's true that they are expanding their scope all the time. Siri currently supports 21 languages in 36 countries, with dialects for Chinese, Dutch, English, French, German, Italian, and Spanish. Alexa supports dialects for English, Spanish and French, along with another 5 languages.

Google added a further 24 languages in 2022, including Quechua, Guarani and Aymara - indigenous languages of the Americas. For the USA, Google Assistant has 10 voices (6 female, 4 male), and English is supported for Google Nest speakers including Google Home in regional versions for many countries, ranging from Belgium, through Australia and Israel. In some countries languages are paired, so that - for example - in India the Assistant can be switched between Hindi and English. Not counting regional variations, particularly of English and Spanish, there are around 20 Google Assistant languages, and the list is growing.

In the voice banking space the industry size is projected to reach $3.7 billion by 2031, growing at a Compound Annual Growth Rate of 14.5% from 2022 to 2031.
Voice banking brings *personalization* to the party, and that's what customers are demanding, particularly in the wake of the global pandemic when many more people started to use and experience relatively seamless online working. Now people want convenient, contactless banking services that speak *their* language, and even their dialect of that language. The idea of struggling with an 'almost-service' is no longer acceptable, as customers know that it's possible to have a reasonably good 'conversation' with a banking bot that will enable them to easily achieve a range of actions. **Increasing seamlessness brings the convincing feeling that banking services are indeed personal, and specifically tailored to the user because they are in essence conversational**.
So let's see who is doing what out in the big wide world, and already putting all this voice work into practice.

NL JP ENG SPA

Flows into practice

USA

Let's start in the USA, the largest market by region for financial voice services. Kathryn Anderson until recently led a Strategy and Design team for Bank of America, with 25 people in the Research and Innovation Lab, itself part of the 150 person Experience Design team.

Is voice now usable and functional? Kathryn reflects that **it's come a long way in a short time**, "When I started, it was very limited in terms of the backend technology. There were a lot of handoffs that would push you to a screen to complete. Now that's shifted a lot." She says that in just a few years almost everything that a BoA customer could do via a screen, can now be done with Erica, the bank's Virtual Assistant, which is 'Personalized and proactive,' and includes many voice functions. Kathryn cites the app as getting more predictive, understanding more of customer needs, then becoming customized to those needs.

"In the earliest form of Erica, it was just, Okay, you ask the right question, and here's the answers we have for you. **Now there's a lot more understanding behind needs, and how customers speak, enabling the more proactive approach.**"

Example? "Let's say you have these subscriptions that you're paying for every month, but are you using them? Kathryn explains. You might be reminded of this by Erica. Voice is getting more and more valid, but there's plenty of room for improvement, and a lot further to go." She sees voice use as a huge opportunity for banking, and bringing the voice experience to customers. So instead of having people come into the app for their banking, or to figure out what's going on with their finances, that information will be brought to where they are. Can their bills be in the calendar that they use every day? Can we get beyond that into doing transactions and using other voice enabled areas?

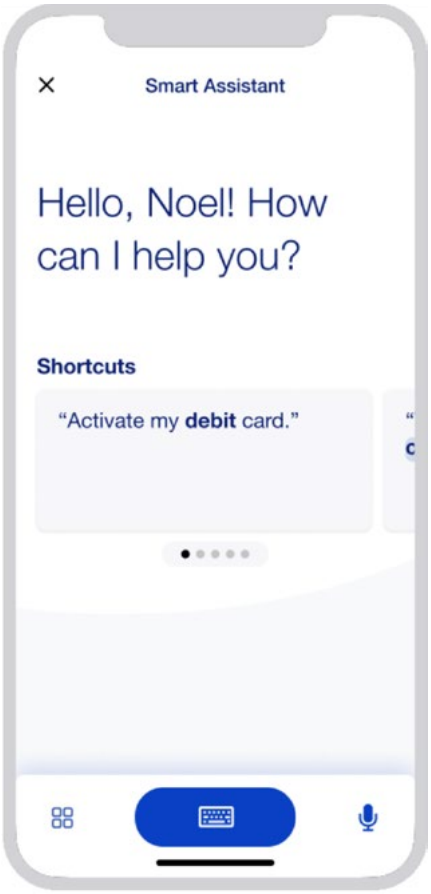
Erica was launched in 2018 and is claimed to be one of the most accessed virtual banking assistants, with around 32 million customers making - as of Spring 2023 - over 1 billion transactions. A feature offered from early 2023, is that the bot can seamlessly hand over to a human Bank of America agent as and when more help is needed. Then when the issue is resolved by the human operator, they can hand the customer back to Erica - so that there is a strong sense of personalization throughout the interaction, all achieved vocally. So far **95% of BoA customers using this hybrid bot/human/bot service have reported that they are satisfied with the new way of working**.

Erica provides a wide range of services, such as monitoring recurring charges and increases, and automatically providing notification of duplicate

charges if they occur. A very useful feature is that when regular payments are due, Erica reminds users that a bill needs to be paid. And when all the bills are paid, the bot provides spending summaries with weekly updates. Plus, for anyone who *thinks* they settled a bill but can't quite trace it, Erica locates past transactions across all accounts. This cross-account tracking is very useful to users, and 'helps you make the most of your money' - on a 24/7 basis, of course. Investing with Merrill can also be achieved through the app, and in addition Erica can access the Zelle digital payments network.

The service is monolingual - in English only - but Bank of America reports that Erica, 'is expected to learn Spanish.' Another near-future intention is to extend from the current mobile-only app to also offer the virtual assistant in desktop banking versions.

U.S. Bank is currently mobile-only with its smart assistant, which can be accessed by typing or by voice and, 'Understands how people actually talk, so you can speak naturally.' This is key to the new breed of highly-trained bots which allow people to use all their many and varied ways of giving commands and asking questions. The bank claims 'a unique voice-first design which progresses beyond the typical chatbot interaction to immerse U.S. Bank customers in an experience of speech response, animation, visualization, and sonic design. The result is a new type of experience for customers unlike anything else in the market for financial smart assistants.'



And what does that claim of 'sonic design' actually mean? Well, it's tied up in the 'voice-first' approach of **the bank where a conversational tone is all-important, and on the bank screen the microphone symbol is dominant, and the keyboard secondary**. Spoken words create a richer data set than just tapping at a keyboard, and the bank's Chief Digital officer Ankit Bhatt has said, "We wanted to leverage voice technology to take our experience to the next level. This includes enabling users to employ nicknames for people (and for their accounts), and using natural language and terms that suit the customer, rather than having to meet the needs of the bot."

As with Bank of America's Erica, the U.S. Bank bot will hand a customer on to a human operator if the situation requires it. Payments, transfers, bill splitting, upcoming bills, spend history and 'enriched transactions' are all possible via voice. (An enriched transaction is where you can ask for a monthly tally of your total spend at a particular coffee shop, for example).

U.S. Bank has a network of some 2,243 full-service offices in 24 states, with over 11 million customers, so it's interesting that despite the density of bricks and mortar branches, there's such a strong push to mobile and voice use. Or looking at it 'through the other end of the telescope' perhaps the number of branches is precisely why the bank has invested so heavily in voice. Will the future see fewer traditional branches? The bank had already closed 257 branches by the start of 2023 so it's a fair guess that - as with many other

banks in the USA and Europe - the drive to cut costs is supported by the adoption of technologies such as voice recognition and activation. Yes, **it's great for customers and personalization, but it's also an increasingly cost effective tool for the banks**.

Wells Fargo launched its virtual voice assistant, named Fargo, towards the end of 2022, with a relatively modest rollout of services starting with users being able to, 'Check credit limits and search for specific transactions by date, amount or type.' Fargo is built on Google Cloud's Dialogflow AI platform and can, 'access a user's bank data including a customer's credit card, checking, savings and lending for home, auto and personal loan accounts.' Upcoming in 2023 is a Spanish language version of Fargo, and the declared intention is to leverage predictive analytics, in a similar way to BoA's Erica. Did you sign up for that special offer subscription all those months ago, and then forget about it? Do you want to continue with it, or would you like to review your spending? The time is rapidly approaching where virtual banking assistants could be initiating the calls to customers to offer new services, such as lending, or connections to the insurance ecosystem. Cold calling from a bot? **The advantages to financial institutions in terms of productivity and reduced staffing levels are hard to ignore**.

Capital One was the first American bank to launch an AI personal finance assistant, way back in 2017. Named *Eno,* it is gender-neutral, and available in both browser and mobile versions, and also communicates across email and text messaging to sports devices such as smartwatches. But is it voice-enabled? No, it is not! I include Eno here however because it demonstrates the stretch between what was clearly an early adopter of AI tech, and the apparent lack of development of voice services. Meanwhile at the other end of the scale is Erica, now increasingly well-established as a go-to virtual voice assistant. Capital One is - by the way - ranked as 10th largest U.S. bank by total assets, with Wells Fargo coming in at 4th, two places behind Bank of America. Do deeper pockets indicate higher likelihood of developing virtual assistants using voice? And, not to be too hard on Capital One, the leader of the pack in the USA, JPMorgan Chase has the Chase Digital Assistant, currently with only screen-based chat available. Yes, the use of voice-based solutions is growing, but not uniformly, although **it is reasonable to assume that behind the scenes most organizations are looking closely at developing an offering**.

Let's now go south of the border and check out if voice is happening in the Latin America region.

Mexico

Banorte in Mexico is one of the four largest banks in the country, and the second largest financial institution overall. It has been operating its virtual banking assistant, called Maya, since 2018, with added voice functions from 2019, saying, 'Maya can assist users through both chat and voice channels. As with any AI-powered tool, the challenge was to achieve a natural, flowing conversation with customers.' In order to create these types of interactions, Banorte used IBM's Watson, where training and calibrating the system only needed to be done once before implementation throughout the bank. 'This speeds up time to market and has allowed the bank to quickly roll out the assistant to all channels.'
The Net Promoter Score (NPS) for Banorte has increased since adopting Maya, with the bot assisting over 47,000 customers a month, and rising. It's the first (and so far only) bank in Mexico to use Google Assistant Natural Voice Technology and Maya can be accessed through mobile and desktop, 'Guiding customers by answering questions about financial products and services; helping with payments, transfers or raising claims; managing account statements; and executing payments and transfers in a more efficient and expedited way.' Amazon Alexa app and Amazon Echo devices are also supported. So, IBM Watson was used to provide the overall training 'brain', while the Google and Amazon assistants provide the voice.

How do Google and IBM work together, and what are the differentiators between Google Cloud AI and IBM Watson? Well, TrustRadius defines it as: 'Google Cloud AI provides modern machine learning services, with pre-trained models and a service to generate tailored models,' while 'IBM Watson Studio enables users to build, run and manage AI models, and optimize decisions at scale across any cloud.'

IBM says that its Watson Assistant, 'Brings conversational AI to your business, and seamlessly connects the channels where your customers engage with the systems, tools and processes that power your business - without migrating your tech stack.' Gartner provides a detailed comparison between Google and IBM Watson, based on over 500 user reviews. However the point with Banorte Bank is that *both* platforms are being used in tandem.

As we are in Mexico, I'll also mention a chat Andras Rung and I had with Alfonso Roibas, Managing Director, Head of Strategy, and Chief of Staff at HSBC Latin America. We were discussing entirely different matters, and when it came to the subject of voice assistants, Alfonso commented that he sees conversational banking as likely to be a game changer in the coming years. **Only 49.1% of Mexican adults ages 18-70 have a bank account, so there are huge opportunities for banking growth, and why would new customers be looking to old tech to get onboarded to the system?** Voice will be a new channel to interact with clients, although Alfonso sounds a note of caution around issues of authentication. On the other hand, deploying call center colleagues to add more value, rather than simply doing jobs which bots can readily undertake looks pretty sensible, going forward. I'm also struck by the fact that with some 126 million people, Mexico has the world's largest Spanish speaking population, and the more mono-lingual a country is, the easier it should be to create voice services.

BBVA - (Banco Bilbao Vizcaya Argentaria) is one of the largest financial institutions in the world with a presence in Spain, South America, Turkey, Italy, and Romania. In Latin America the Blue conversational voice assistant is available in Mexico, Colombia

and Argentina. The project began in 2018 with the challenge of bringing together various existing bots across the group, and producing a new, more fit-for-purpose app. Starting with a relatively small team, the task was to 'research, compile, retrieve, and redefine the principles, guidelines, tools, and assets. As well as the original two UX-UI designers, and a conversational designer, the team grew to include twelve designers, six product owners, three program managers, and twenty developers across the participating countries - a global endeavor.'

As pointed out at the beginning of this article, getting the 'personality' of a bot is important, and this was examined in detail for Blue, with initial user research in four countries (Including Spain) being followed by, 'an arduous analysis process, which included psychology and positioning methods. It was important to be consistent, ethical, and responsible in creating the tone our virtual assistant would use to communicate with customers.' OK, so what *is* the tone of Blue? 'Blue is pleasant, patient, and thorough. When it comes time to work, it is attentive and sleek. During more relaxed times, it likes to show off its wit.'

Aha - Now *that* sounds like 'conversational'! **Is the time coming where we'll be calling our digital banking assistant to have a bit of a chat and tell a few jokes?**

Canada

If we head north from Latin America, leapfrogging the USA, we come to dual-language Canada, where credit union provider Central 1 launched the country's first authenticated voice banking service in 2018. Using Amazon Alexa, 'Customers can make payments, send money to vendors, transfer money between accounts, and better understand their financial wellness with just their voice. Central 1's smart, conversational user interface pairs hands-free functionality with human-like dialogue interaction, and leverages machine learning and artificial intelligence to continually improve its own accuracy.' Interestingly, given our remit of all things Fintech, Central 1's *Forge Community* is a way of connecting contributors and participants in digital banking solutions, performing like an incubator-stroke-innovation hub.

Tangerine, a subsidiary of Scotiabank, which ranks as Canada's third largest, has had voice banking since 2018, when it was described as 'Taking mobile banking options to a whole new high-tech level.' Based on the Nuance conversational AI platform, Tangerine offers true Voice Activation (as opposed to the Voice Recognition spoken of earlier) enabling users to access services biometrically with their own pre-stored and unique voiceprint. Nuance was acquired by Microsoft in March 2022 and is focussed on healthcare and medical diagnostics, as well as, 'helping organizations in every industry create more personalized and meaningful customer experiences', with the goal being that, 'Microsoft and Nuance will enable organizations across industries to accelerate their business goals with security-focused, cloud-based solutions infused with powerful, vertically optimized AI.'

And from Canada, let's travel east to Europe… to Central Europe in fact.

Poland

mBank in Poland is the country's fifth largest bank in terms of assets, and was the first to offer mobile services, starting in 2014. The bank itself is a relative newcomer to the Polish financial scene, having been set up in 1986 as *Bank Rozwoju Eksportu,* the Export Development Bank. Marek is the bank's conversational voicebot (and yes, he's a *he*), allowing a variety of 'basic' activities such as setting a card PIN, activating a card, or changing a card transaction limit. Marek will also hand over a customer to a human consultant if the process requires it. mBank gives some helpful tips to customers thinking about engaging with Marek, which may just as well apply to all interactions with voice applications: 'How to chat with a bot? - Naturally, like a human! Try to speak in simple sentences and do not interrupt the bot's speech. If you don't hear or understand something, you can always ask the bot to repeat it.' Marek is available in Polish only, for a population of around 38 million Polish speakers, of whom mBank has 3 million active users of mobile banking.

At Alior Bank, all calls are currently answered by the virtual assistant infoNina, which directs callers to consultants and handles up to 10% of cases without having to switch the connection. The bank says that **around 42 million public statements are analyzed every day to expand infoNina's knowledge of natural language, including different accents, colloquial speech… and even profanities!** With constant training of the algorithms, Alior makes a point of noting that the bot is a *work in progress*.

When in doubt, infoNina will hand on to a customer service representative, as is the case with the IKO voice assistant - using Siri - from PKO Bank Polski. IKO operates through text and touchscreen in Polish, Ukrainian, Russian and English, and in Polish is able to help with 160 different topics of conversation. With over 11 million Ukrainian refugees fleeing their homeland through Poland, and over 1.5 million staying there, having IKO available in Ukrainian must help to keep displaced people financially linked to those remaining 'back home'. Indeed there are now also many Ukrainian entrepreneurs setting up new companies in their host country: According to the Polish Economic Institute, nearly 14,000 Ukrainian companies were established in Poland between January and September 2022. Developments in 2023 are promised to make the IKO bot understand over twenty new types of instruction, as well as **being tailored more towards helping blind and partially-sighted people -** a service which András Rung of Ergomania has underlined as being an important feature for the future of voice.

Italy

Southwest to Italy and the Fintech company Auriga, of 'The banking e-volution' which, 'specializes in software solutions in the banking area, with the development of integrated platforms that have contributed in defining the standards of the new models of omnichannel banking.' The IOLE virtual assistant offers fully automated help to customers of Banca Carige (now BPER Banca) through the Bank4Me platform, which Auriga claims reduces conventional operating costs by 38%. IOLE is a ' powerful virtual assistant that intuitively and fully answers customers' needs and provides extended assistance.' And, as I'm sure you'll know, *Iole* was the name of the beautiful daughter of King Eurytus in Greek mythology!

Spain

BankInter has BIA and Beatriz as its virtual assistants, both developed with IBM. Beatriz is for the Spanish-speaking market, and BIA (BankInter Interactive Assistant) for Portugal only. Both bots perform the 'usual functions' and will hand over to a human operator as and when needed.

Founded in 2012, EVO Banco is one of Spain's largest digital banks, with over 900,000 customers. The bank's EVO assistant is the country's first AI assisted voice banking app which engages with customers in a 'hyper personalized way' - in Spanish of course.

EVO includes biometric ID by repeating the stored phrase, 'In EVO, my voice is my password.' This, to some extent, is a 'hostage to fortune' as debate continues about the security of voice login, especially as deepfakes and generative voice AI systems are increasingly sophisticated and common. *In theory* voice verification is more secure than passwords because it's harder for hackers to steal your voiceprint or biometrics data, whereas a password can be stolen by anyone who sees it, or hacks it. 'Liveness detection' is one tool seeking to prevent misuse of voice authentication, by using prompts to assess how a user responds to questions.

But back to Spain. EVO Banco has been using Google Cloud services since 2015, so the platform was a natural choice, as EVO Banco's Head of Disruptive Innovation, Big Data and Advanced Analytics Pedro Tomé says, "We needed something reliable and quick that could respond to human questions just as naturally as we are talking now. We didn't want a chatbot, we wanted the conversation to flow naturally as in real life. **We needed a system that could connect with all the technology in our roadmap,**

that we could develop, scale up, and flex with demand, all while optimizing costs."

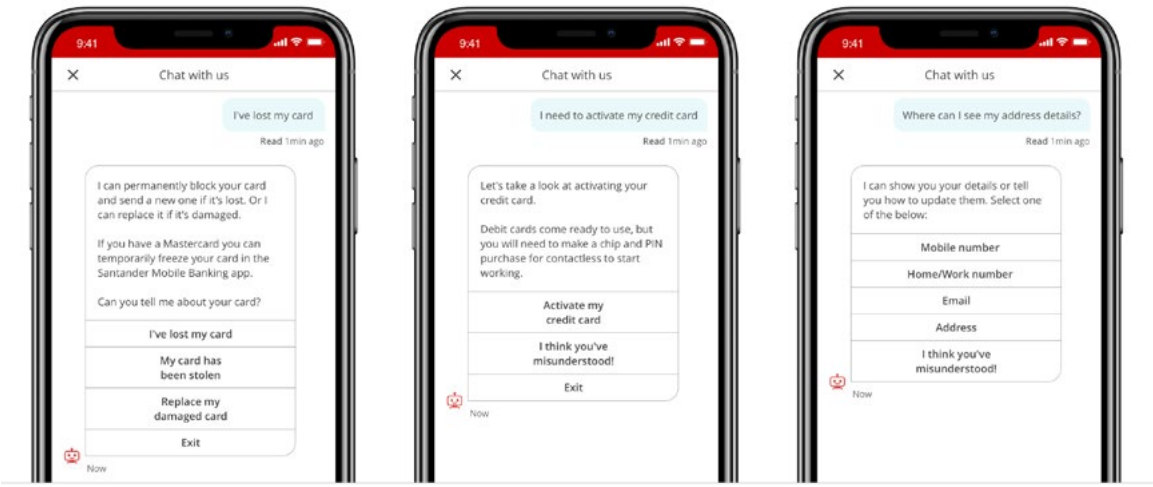
Around 85% of calls to the contact center are now handled by Dialogflow, and the average time people spend on the line for an answer is two minutes, where it was previously five to six minutes. In that time a customer has called, had a conversation with the system, got what they wanted and hung up. The accuracy of correctly-routed calls is said to be 95%, and the 'conversational effectiveness' - yes, that's a real metric - is at 70%. While that sounds relatively high, it does of course mean that **30% of interactions between customer and bot are not reaching a successful conclusion**. Typically this will involve the voice assistant using phrases such as, 'I'm sorry, I didn't understand you, please repeat,' or the handover to a human operator. That's not a critique of EVO Banco's offering, simply the observation that while great strides are being made globally in financial voice assistants, there is still a way to go.

UK

While Banco Santander originated in Spain, it's in the UK that it became the first to launch voice banking. "We believe technology, like voice banking, will play a transformational role in the way we add value by creating more choice and convenience for our customers. Voice is the most natural way we interact, and with a growing number of customers using mobile banking, Santander has merged the two experiences to create a quick, simple and secure hands-free option." So said Nathan Bostock, Chief Executive Officer of Santander UK, back in 2016. By 2017 the SmartBank app capabilities were being extended, on the Nuance platform, with specific attention to the needs of more vulnerable customers, including blind people, and those unable to easily leave their homes.

The bot, now named Sandi, can undertake all the 'usual' tasks such as reporting lost or stolen cards, activating a card, or updating personal details. Users are advised, **'Keep it simple. Try to ask one question at a time and keep sentences short and easy to understand:** For example, 'I want to order a new card'.

It seems that it will still be a little time before we can be *truly* conversational with voicebots, and use our full range of annoying human vocal habits, with all the tics, ums, ahs and slang!

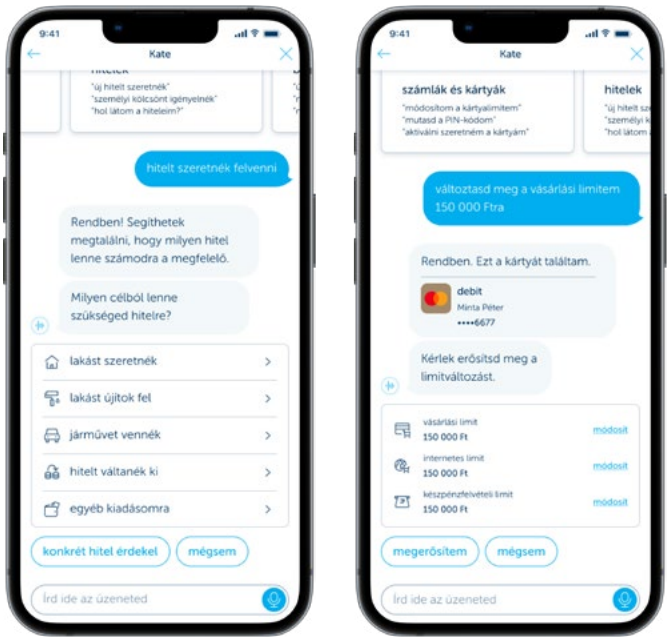


Benelux, and beyond

The KBC Group is a Belgian multi-channel bank-insurer, focusing on private clients and small and medium-sized enterprises in Belgium, Bulgaria, Czechia, Hungary, and Slovakia. In 2020 KBC launched its Kate personal digital assistant in Belgium, along with the help of Ergomania Digital Product Design for the roll-out of the Hungarian version in 2022. From the beginning the concept for Kate was to support text chat and voice chat, driven by Rasa AI. In an interview in TheBankingScene, an interesting observation made by Katrien Dewijngaert, General Manager Strategy and Transformation at KBC Insurance back in 2021, is that in early evaluations, users found the system 'non-judgemental'. In other words, while a human consultant might react to - for example - a customer going over their credit limit, Kate takes it in her stride, without any tonal suggestions of disapproval. **People like, Knowing that it's not a person, it's not somebody who can judge you.**

Renáta Szilágyi-Nagy picks up on this, and how the personality of Kate was a key factor from the start of the project. It's the same Kate wherever you go, Renáta says. She's personal and warm, speaking to users in an informal way. I question that 'informal', because Hungarian, for instance, has both formal and informal ways of addressing people. How was that dealt with? Although the Hungarian bank typically used the formal address in its normal communications, for the voice project, the team were adamant that informal had to be the way forward. This was mainly about 'tone', and keeping things friendly sounding, but there were also the practical considerations of not doubling up on language for the formal/informal grammar. Were users initially troubled by this informal approach? Renáta reports that the general response in user testing was, 'I don't mind the informality... but I do know other people who might.' So far however there has been little pushback from customers - a mark of the increasing informality in general society, and indeed the trend of many people to intentionally avoid formal speech.

The Ergomania team needed to change Kate to become Hungarian, which was essentially a large-scale translation job... or not? Renáta says no, it was much more than substituting one language for another. **The *personality* of Kate had to be consistent, but the language used needed to reflect local conditions**. Renáta gives as an example terms which in translation can be interpreted inaccurately, such as the differences between 'Stop', 'Decline' and 'Cancel'. What works in one language cannot simply be *copy pasted* across to another, and will need to be differentiated, or in some cases even combined. Language-wise, similar terms can confuse a system, Renáta explains, So there's always reworking to be done. It's an almost endless task, because the bigger and smarter Kate gets, the more use cases there are, and the larger the team needed to 'feed' the bot. Back-end processes are different in each coun-



http://kh.hu/kate

try that Kate works in, so there can also be a lot of below the surface activity needed to adjust to local conditions. An instance of this is the relatively simple activity of money transfers. Belgium legislation and practices are different to Hungarian, so again, *copy paste* solutions won't make the cut.

I return to the remark made by Katrien Dewijngaert about Kate being 'non-judgemental'. Is that the case with the Hungarian Kate too? Renáta agrees, and adds that Kate is, "Never pushy. She talks at a human level and *advises*. She is not a sales or marketing machine. **It's customer-focussed communication, and helping as much as possible from the user's point of view.**"

And what about the issues of privacy, confidentiality, and security? I imagine that having such a friendly and natural-sounding bot to chat with could result in my details being overheard by everyone on the bus. Not so, says Renáta, Kate is designed to offer to *show* onscreen any sensitive information, such as a new pincode, or account balance. The voice aspect is always active, but the bot will check if it's secure to speak aloud about important confidential details.

Kate can now help find what customers are looking for within the bank, and assist with all kinds of banking and insurance matters. These include transferring money, or ordering foreign currency. Insurance claims are assisted by Kate, starting with reporting of a car accident through the app. That's a really useful feature for when a customer may be already stressed, and not able to put their hands on the right paperwork. Kate files the claim and puts the customer in touch with their insurance agent. Hospital admissions also come under Kate's remit, and the bot can start the process of filing a health insurance claim.

There are also a host of functions around debit and credit cards including help for a forgotten PIN, and the ability to block a stolen card through Kate, or request new cards. And of course Kate will give customers their billing statement, and even have a chat about increasing a credit card limit.

And the name 'Kate' by the way actually has a meaning, although I doubt you could easily guess at it…

OK, it's **K**BC **A**ssistant **T**o **E**ase (your life). Oh, you knew that?

UAE and Middle East

Jumping over 5,000 kilometers from Belgium and Hungary, we come to the United Arab Emirates, where the Dubai-based and government-owned Emirates NDB bank has a voice assistant utilizing Amazon's Echo in Arabic. English is also supported (and can be translated), along with Khaleeji, the local Gulf Arabic dialect. The voice assistant, named EVA, was introduced in a pilot scheme as far back as 2016 with the intention of, 'enabling quicker, simpler, personalized and intuitive customer service.'

There are around 420 million Arabic speaking potential customers in the region, but despite the advances with EVA, the language is still tricky to use in the sphere of AI. This is because the sounds and grammar of Arabic can generate many different words (and therefore meanings) from the same 'root'. Capital letters aren't used with proper nouns, which make it more difficult for AI systems to detect (for example) the *person* in a sentence using 'named entity recognition'. So training a system be-

comes much more complex in Arabic, including the multiplicity of Arabic dialects. For example the Rafiq personal assistant allows phone commands and responses with a Saudi or Egyptian accent, but has very limited functions, such as hailing a taxi ride.

The processes behind a voice assistant revolve around converting speech to a text document, which is then 'understood' by the system using Natural Language Processing. The process then runs in reverse, with the resulting text answers being converted back into human-sounding conversational speech. Layering in the complexities of meaning, and many different dialects across the region mean that **as yet there is a lot of activity in Middle Eastern voice banking, but fewer results, precisely because of the lack of precision in the available bots**. For example, Egyptian Arabic is claimed to be one of the fastest spoken dialects in the world, making it hard for any existing bot to hang onto every word. Why? Because the user's spoken word is first captured as text, and there is a finite speed that this can be done at... By which time the next line has been spoken, so there is a constant buffering, coupled with the search for optional meanings. This optional meanings issue is key because there is, as yet, a significant lack of large scale Arabic datasets to draw from, making it more difficult to train accurate models. So in short, the system is constantly playing 'catch up'. There are however English and 'Arabic Dialectal' bots from the likes of Xina in Jordan, and Nvidia's Jarvis Arabic voice assistant making inroads into the market.

And that said, the Israeli Bank Hapoalim (in Hebrew and Arabic), Discount Bank, and Leumi Bank have all had operational voice biometric technology since around 2010.

India

According to a 2022 PwC report into voice technology in India, **voice search inquiries are growing at 270% a year, and growth in smartphone usage is set to make India the largest voice-first market**. This is helped in part by the rapid uptake of phones costing the equivalent of $20, which have a dedicated voice assistant button. **This is leading to a 'voice everywhere' culture which can only help support voice banking initiatives**. The overall literacy rate in India, according to the National Statistical Commission was 77.7% in 2011, meaning that nearly a quarter of the population are limited in their use of screen and text-based apps. So voice use is set to continue growing, especially when coupled with access to cheap mobile phones.

Alexa is now in Hindi, but Cortana is not, while Siri works in Hindi, Telugu, Kannada, Tamil, Bengali, Marathi, Punjabi, Malayalam, and Gujarati. This, in a country with 120 major languages, and in excess of 1,500 dialects. In banking, a 2022 study by Cognizant showed that 94% of banking and financial services executives surveyed in India believe that the shift towards voice will only accelerate in the future, and **72% said that voice is important, or extremely important, for their bank's future success**.
2020 saw the launch of the iPal bot using Alexa and Google Assistant, for ICICI Bank, one of India's leading private sector banks. Axis Bank also has its offering, AHA!, an automated voice assistant in Hindi or English that can recognise the intent and nature

of customers' queries, and respond appropriately and effectively.

Voice banking is being actively used in India, and the economics and demographics, plus the intentions of banks, point to significant developments in the very near future.

Bangladesh

Bank Asia is a private sector commercial bank dating back to 1999 with a big footprint of ATMs and bricks and mortar branches across the country. Biometric ID, in the form of fingerprint recognition, was already in use, when at the end of 2021 its voice assistant was introduced, in Bangla and English as well as the local dialects of Noakhali, Sylhet and Chattogram.

...and Everywhere

We could also go to Turkey, Sweden, Germany, or well, almost anywhere, and there will be developments in voice banking. Although interestingly in that bastion of world banking, Switzerland, a recent chat suggested that voice is *not* yet the hottest of topics. Martin von Siebenthal is Circle Lead at UX design agency Ginetta, and remarked that with Switzerland being a four language country, there are *opportunities* for translation apps (for instance), but which are also therefore made more complex. "The challenge in designing voice interfaces is, what would someone say at this point? There are all these dialects too, and maybe it's also a cultural and generational thing, in that voice is less used in Switzerland. Until recently text messages were the predominant means of connecting between individuals."

Top users of voice banking worldwide include:

Bank	Platform	Language	Features
Europe			
ING Netherlands	Nuance	Dutch	Nearest branch, payments, Voice Biometrics (Currently non-operational)
N26	Siri	English/ German	Main Account breakdown, Spaces overview
OTP Bank	Nuance	Hungarian	(Currently non-operational)
Volksbank	Alexa	German	No information currently available
Bundesverband des Deutschen	Alexa	German	n/a
Die Sparkasse Bremen	Alexa	German	No information currently available
Die Sparkasse Bremen	Custom	German	Voice bot in customer service
Sparkasse Marburg-Biedenkopf	Alexa	German	No information currently available (Opening times, ATM location)
VR Bayern Mitte	Alexa	German	No information currently available
HypoVereinsbank	Alexa	German	No information currently available
Volksbank Raiffeisenbank Laupheim-Illertal	Alexa	German	No information currently available
Volksbank Uelzen-Salzwedel	Alexa	German	No information currently available
Volksbank RheinAhrEifel	Alexa	German	No information currently available
Bank Austria	Alexa	German	No information currently available
Bonify	Alexa	German	n/a
Sparkasse	Google Home	German	Balance, info about payments
Sparkasse	Siri	German	Send money

Bank	Platform	Language	Features
Deutsche Bank	Siri	German	Send money
Postbank	Siri	German	Send money
BW Bank	Siri	German	n/a
KBC Belgium	Siri	n/a	Multiple banking functions
UBI Banca	n/a	Italian	n/a
Sella	Google Home	Italian	Money transactions
Webank	n/a	Italian	n/a
Rabobank	Google Home	Dutch	Request your balance and latest transactions Set up a daily update for your checking account Share your last bill or make a payment request Check your maximum mortgage
DNB	Google Home	Norwegian	n/a
SpareBank	Google Home	Norwegian	n/a
SpareBank	Siri	Norwegian	Transfer money between your accounts
Skandinaviska Enskilda Banken	Siri	Swedish	Providing information and services through Google Assistant
Bank Millennium	n/a	Polish	n/a
Millennium BCP (Banco Comercial Português)	n/a	Portugal	n/a
Abanca	Siri	Spanish	Send money
CaxiaBank	Text chat with voice	Spanish	n/a
CaixaBank	Google Home	Spanish	No information currently available
CaixaBank	Alexa	Spanish	No information currently available

Bank	Platform	Language	Features
CaixaBank	Siri	Spanish	Send Money
BBVA	Siri	Spanish	Send Money
BBVA	Alexa	Spanish	n/a
Bankia	Alexa	Spanish	No information currently available
Bankia	Google Assistant	Spanish	Use Google assistant to boost customer service
Banco Santander	Alexa	Spanish	n/a
ING Poland	Google Assistant	Polish	Checking balance, ordering a transfer, generating a Blik code
K&H	n/a	Hungarian	Kate - Account information
Wells Fargo	Google	English, Spanish	Fargo - Bank balance, credit card balance, money transfer
North America			
PayPal US	Siri	English	Send money
Capital One US	Alexa	English	Account enquiries, track spending, payments, balance, pay bills
Monzo	Alexa	English	Account enquiries, payments
Monzo	Siri	English	Balance
American Express US	Alexa	English	Check balance, make payments, get Amex offers
US Bank	Alexa	English	Account enquiries, track spending, transfer between your accounts
USAA	Alexa	English	Account enquiries, payments
Enrichment CU US	Alexa	English	Account enquiries, payments, loan payments, text or email balances
Ally Bank US	Alexa	English	Account enquiries, track spending, payments, 'CurrentSee' feature, transfer money

Bank	Platform	Language	Features
Bank of America US	Siri	English	Balance info, reminders etc.
Wells Fargo	n/a	English	Voice verification
Wells Fargo	Siri	English	n/a
Royal Bank of Canada	Siri	English	Send Money, Pay Bills
Town and Country Federal Union	Alexa	English	Balance info
Virginia Credit Union	Alexa	English	n/a
American Express	Alexa	English	Make Payment, Balance info
Starbank	n/a	English	n/a
D3 Banking	Alexa	English	n/a
First Hawaiian Bank	Alexa	English	Balance info
Nation Trust Bank	Alexa	English	n/a
Nation Trust Bank	n/a	English	Voice verification
Summit Credit Union	Alexa	English	Check balance, transfer money
Enterprise bank	Alexa	English	n/a
Numerica Credit Union	Alexa	English	Check balance, transfer to savings account, pay auto loan
J.P. Morgan	n/a	English	Chase - Send money, account information
U.S. Bank	n/a	English	Smart assistant - Send money, account information
Venmo	Siri	English	Send money
UK			
Barclays UK	Siri	English	Make payments
Barclays UK	n/a	English	Voice biometrics in phone call

Bank	Platform	Language	Features
Starling Bank UK	Google Home	English	Account enquiries, payments
Santander UK	Nuance	English	Payments
Santander UK	n/a	English	Voice biometrics/verification
ebankit	Cortana	English	n/a
Mercantile Bank	Siri	English	Transfer money
Asia Pacific			
KEB Hana Bank Korea	Bixby	Korean	Account enquiries, track spending, financial news, payments
Shinhan Bank Korea	Bixby	Korean	Account enquiries, track spending, financial news, payments
Woori Bank Korea	Bixby	Korean	Account enquiries, track spending, financial news, payments
Woori Bank Korea	Bixby	Korean	Voice phising detection app
ICICI India	Siri	English	Send Money
ING Australia	Siri	English	Account enquiries
Ant Financial Group	Unique development	n/a	Purchasing tickets, booking hotels, calculating daily returns, balance checking
NAB	Alexa	English	Account enquiries
Westpac Australia	Alexa/Siri/ Google	English	Account enquities, financial news
OCBC Singapore	Siri	English	Bank balance, credit card balance, money transfer
HDFC Bank India	Alexa	English	Basic bank information
CIMB Malaysia	Unique development	n/a	Account enquiries, track spending, payments
Garanti Bank (BBVA)	Unique development	Turkish	Transfer money, basic info, buy/sell foreign currency

Bank	Platform	Language	Features
Akbank	Siri	English	Pay bills, Transfer money
Shumishin SBI Net Bank	Alexa	Japaneese	Basic info, current balance
MEA			
Emirates NBD UAE	Unique development	n/a	At call centres intelligent call routing and account enquiries, English and Arabic
Emirates NBD UAE	Alexa	English	Account information
South America			
Bradesco	Alexa	Portuguese	Send money, account information
Itaú	Alexa	Portuguese	Information, discounts
Banco do Brasil	Alexa	Portuguese	n/a

The trends in voice

So we've dotted around the globe, somewhat randomly I admit, and seen some trends in voice use for banking and financial applications. What conclusions can be drawn? Here's my stab at making some sense out of all the product launches, reports and reviews:

- **Voice banking is on the cusp of becoming ubiquitous,** and my guess is that there's hardly a bank on the planet that is not interested in developing solutions. Why?

- Because **voice assistants work 24/7** and are increasingly able to handle complex queries.

- When they can't meet a customer's requirements, many voice bots can now **hand over seamlessly to a human advisor**. What's more, the human can hand back to the bot to pick up where it left off.

- Territories which are **largely monolingual** tend to have more **highly developed offerings**, and this is particularly noticeable for the USA, and Spanish-speaking countries.

- This ties in too with the language capabilities of the various speech platforms such as Siri, Alexa, Cortana and Google. Some languages and dialects are well catered for, while others are currently underserved.

- The **demographic and culture** of a country also plays into the **acceptance of voice services in banking**: There may be an inbuilt conservatism present, as in Switzerland, or a massive market for new adopters of smartphones, as there is in India.

- Voice recognition and voice activation also serve the less-abled, such as blind people, as well as anyone who either through choice or work requirements needs **hands free operation**.

- Following the pandemic, touch free solutions are in demand, as **touchscreens and biometric ID** through fingerprints have become less attractive for hygiene reasons, while voice biometrics are increasingly appreciated by users.

- Socially there is less reserve among users about using voice to connect to their bank. It's become normalized, with *Digital Natives* leading the way.

- And when you look at banks, from Bangladesh to Belgium, the simple fact is that **voice assistants are more cost-effective**.

- As many banks - particularly in the USA and across Europe - seek to reduce their bricks and mortar footprint, **voice assistants** close the gap for customers, **offering more convenient access for many services**.

Tools from the toolbox

So we've visited the theory of language modeling, and flows, and seen how these are sculpted into massive datasets used to train AI systems. It's more than 'just' about advanced linguistic theory and practice however. Voice User Interface technologies also have to be constantly reviewed and kept abreast of. So let's return to Ergomania, with the team adept in the use and deployment of a number of leading tools, including:

Nuance Mix, which is promoted as, 'Empowering organizations to create advanced conversational experiences for IVR (Interactive Voice Response automated phone technology), and chatbots. Nuance uses market-leading tools for design, speech, natural language, dialog, and testing. With one tooling platform across the full software development lifecycle, enterprises gain greater control, accelerated development time, and increased business agility. The result is conversational AI that drives business outcomes.'

The software currently enables conversational experience in 86 languages, within the same project: an important feature in reducing - or even eliminating - the need to repeat the same work for each channel. Nuance grew out of Lernout and Hauspie Speech Products, a Belgian speech recognition company founded in the late 1980s by pioneers Jo Lernout and Pol Hauspie. Although the founders were later found guilty of fraud, when the company was declared bankrupt the technology was acquired by Scansoft - later renamed as Nuance.

Similar to Nuance, Dialogflow from Google is a, 'Natural language understanding platform that makes it easy to design and integrate a conversational user interface into your mobile app, web application, device, bot, or IVR system. Using Dialogflow, you can provide new and engaging ways for users to interact with your product.' Dialogflow is able to analyze more than one form of input and can accept multiple types, including text, as well as audio from a phone, or a voice recording. It can then respond to customers both through text, or with synthetic speech. Dialogflow was improved and updated in 2020, with - among other features - a tenfold increase in the number of 'intents' available to virtual agents. The system currently supports speech-to-text capabilities in 32 languages and text-to-speech capabilities in 28 languages, while Google Translate supports translation between 104 languages.

In the roster of software there's also Watson from IBM, one of the **most widely recognized tools in the AI toolbox**. IBM says, 'Watson helps you unlock the value of your data in entirely new, profound ways, giving every member of your team the power of AI. With a suite of pre-built applications and tools to build, run and manage your AI, Watson gives you insights to predict and shape outcomes and infuse intelligence into your workflows.'

Well it's hard to argue with that. The Watson assistant service supports English and 13 other languages in the GA (Generally Available) category, with other languages partially supported, and of course in continuous development.

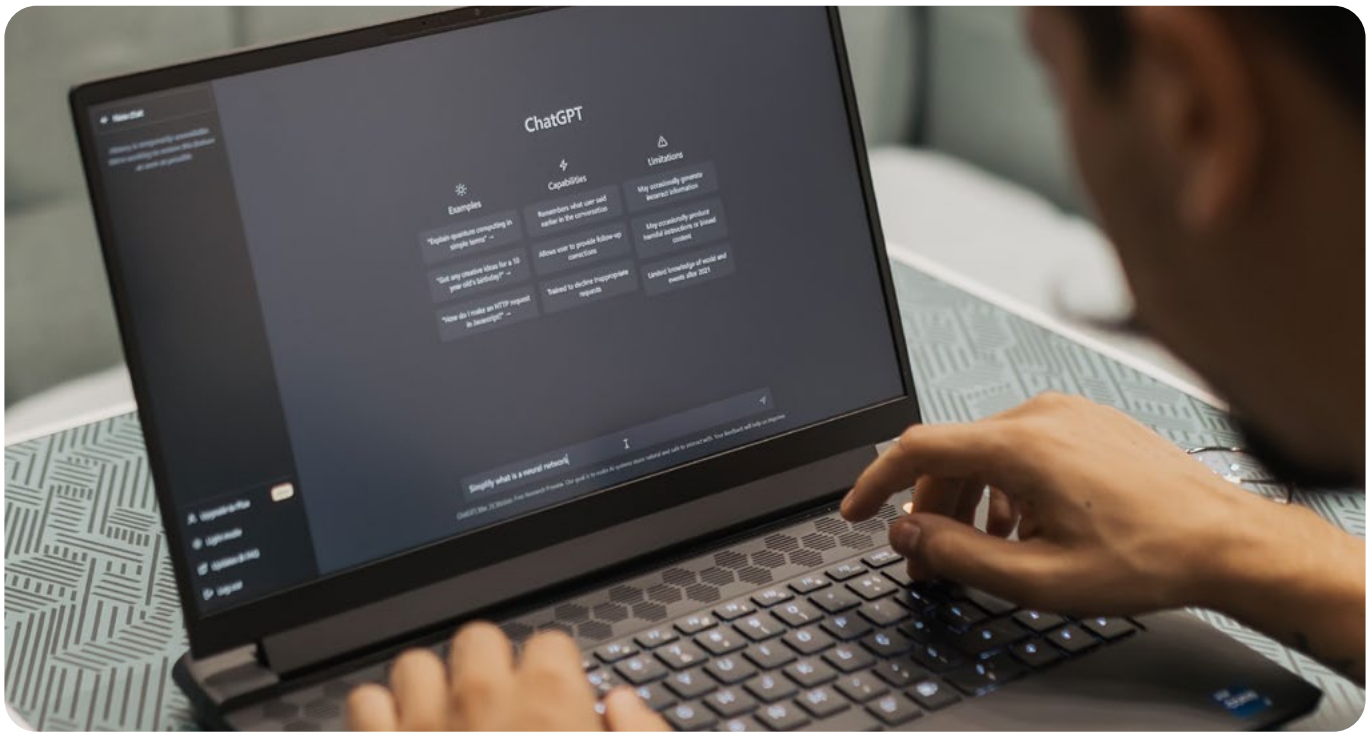
Language Understanding in Action, or LUIS from Microsoft, 'Interprets user goals (intents) and distills valuable information from sentences (entities), for a high quality, nuanced language model. LUIS integrates seamlessly with the Azure Bot Service, making it easy to create a sophisticated bot.' LUIS currently supports 20 languages worldwide, although some of these - such as Japanese - are 'key phrase only' categories.

Challenges going forward

With the sophisticated development tools available, and global banking markets showing ever greater interest in the potential of voice banking, it would look like the way is clear for widespread adoption. But there are still some bumps in the road ahead:

- The first challenge is accuracy. Obviously voice recognition and voice activation technologies must be able to **accurately recognize** and **respond to voice commands** in order to be effective.
- This can be difficult, as voice technologies sometimes **can't differentiate between similar words** and phrases.
- True voice authentication could replace facial and fingerprint ID as being more convenient for users, but this is **challenging in the era of deepfake software**.
- Some languages - such as Arabic - are more challenging than others, such as global English.
- Return On Investment can initially look steep with significant outgoings on development, hardware and software, usually some time before returns are seen in cost-effectiveness.
- There can be no 'one-size fits all' solution for banks - because of regulatory and competitive reasons - so **each organization has to commit to an individual roadmap with dedicated resources**, from either within or without the company.
- Scalability can be an issue as **customer numbers and the complexity of their needs can vary significantly**. In very large and rapidly expanding markets such as India, development at scale for voice will be challenging.

The elephant in the (chat) room



ChatGPT from OpenAI started hitting the headlines at the beginning of 2023. The media shouted about how the software would rescue lazy students from essay deadlines, and journalists wondered if they would soon be replaced by the newest and smartest of AI programs. Since then, we've seen the advent of the even smarter GPT-4, which, 'can solve difficult problems with greater accuracy, thanks to its broader general knowledge and problem solving abilities.'

In March '23 the Irish/American Fintech Stripe began integrating GPT-4 into its digital payment processing - one of the first known uses of this latest AI iteration - and among the fourteen GPT-4 prototypes is the ability of customers to query their own analytics using natural language. Meanwhile there has been a pile-on of ChatGPT-alikes using AI and Large Language Models. These include Bard from Google, AI Bing, TruthGPT ('Discover the unbiased truth to any question that comes to mind!') Socratic, for students, ChatSonic ('The ChatGPT alternative built with superpowers'), and Meta AI's LLaMa ('A foundational, 65 billion parameter large language model').

A noticeable trend in the rapidly developing field of Generative Pre-Trained Transformers is that **banks which seemed proud of their trailblazing use of ChatGPT at the beginning of 2023 have now gone rather coy about it**. Instead of the initial press releases trumpeting actual banking uses, we're now back to *potential* benefits and generic descriptions, such as 'ChatGPT can assist banks in identifying and managing potential risks by analyzing vast amounts of data and identifying potential risk factors. Banks can also use ChatGPT to monitor transaction activities, flag suspicious transactions, and identify potential fraud.' This caution comes in the wake of the Italian government banning ChatGPT in March 2023, and then unbanning it a month later. Meanwhile EU regulators, and governments around the world are weighing the pros and cons of AI and GPT platforms.

So how does ChatGPT work, and what is it doing? We'll use the program as a general example for the flood of 'me-toos' that have followed. For a start it is 'just' a chatbot, which can answer questions, draft emails, converse, translate and explain code, based on text. As OpenAI says, 'ChatGPT works by analyzing the user's input and using it to generate a response based on its vast knowledge repository. **One of its standout features is its ability to remember previous interactions and incorporate that context into subsequent responses.'**

Let's look a little deeper into what's going on 'underneath the hood' of ChatGPT. The system works using OpenAI's GPT 3.5 architecture, which **collects billions of instances of publicly available text from across the web**. It is not looking at *content* but at how language is used - this is the **Pre-Training phase**. Here the model is searching for the patterns of language, syntax and grammar, from which it develops 'understanding' of the probabilities of the order of words in a sentence, and the most-likely next word. For example, in my earlier example, 'There's been an accident…' ChatGPT will capture the fact that the next phrase will most commonly be 'call a Doctor', or a second probability of 'call an ambulance'. It will find few, or no examples of sentences such as 'There's been an accident, call a banana'. Using Deep Learning there is also some reasoning brought into the analysis: My '…call a banana' is illogical, and therefore highly unlikely.

So, based on the context of previous words, the model comes up with most-likely next words, using its neural network architecture consisting of multiple layers of *self-attention mechanisms* and *feed-forward neural networks.* Even the creators of ChatGPT are reported to have been surprised by how this 'basic' approach generates such high quality responses, although this is actually very like real human conversation, in that we often don't know the exact word we will say three words later. So ChatGPT is constantly checking itself against the massive raw database that it was pre-trained on. It does this *without* knowledge of which documents it is taking sentence structures from - they can be literally anything which is available 'out there' in the form of books, websites, and articles of every sort. It is indeed a 'Large Language Model'.

The point of the exercise is to enable the generation of human-like responses and interactions using Natural Language Processing. This can mean answering questions, but equally, creating entirely new text sequences: one of the reasons that academics have been concerned about lazy students generating essays, and journalists and writers worry about their future employment.

At this stage what the model has been doing is '**Unsupervised Learning**' where it makes the best sense it can of the data it has ingested, so it wouldn't offer much of a challenge to professional writers. But training is two-stage, because now comes the Fine-tuning phase with the use of RLHF - Reinforcement Learning from Human Feedback. Now more specific datasets are overseen using human-generated and augmented text which is more natural and 'usable'. This part of the ever-iterative training loop has actual human trainers who score generated texts for accuracy, and feed in language which is specific to a use case.

For example, a banking application which uses ChatGPT as the basis of its bot will require very careful human supervision and fine-tuned training in banking terms and language. Using highly-skilled human overseers (who work to tightly-specified instructions) the model's parameters are steadily and iteratively adjusted to ensure performance optimization for any given task. These parameters are defined by the client team, with the intention of creating the most useful, clear and *truthful* responses.

Truthful? Well, one current observed problem with Large Language Model systems is that there is the potential to continue and sometimes build up biases, or to even 'hallucinate' facts, which are then

perpetuated within the language model. It is probably for such reasons that, in the early months of 2023, banks became more careful about declaring their allegiance to GPT models, after initial enthusiasm.

So what can Chat GPT do for banking? Here's what the software says about *itself* when queried:

'Chat GPT and other advanced natural language processing (NLP) technologies can have a significant impact on the banking and finance industry in several ways:

- Customer Service: Chat GPT can be used to automate customer service functions, providing customers with **conversational 24/7 access to information and assistance, reducing wait times, and improving the customer experience**.
- Financial Planning and Advice: Chat GPT can provide personalized financial planning and investment advice by analyzing individual financial data and goals, and providing tailored recommendations, using voice as the medium.'

The chatbot goes on to add other features and benefits:

- 'Fraud Detection: Chat GPT can analyze large amounts of financial data to identify patterns and anomalies, potentially detecting fraudulent activity more quickly and accurately than manual methods.
- Risk Management: Chat GPT can assist with risk management by analyzing large amounts of data and providing insights into market trends and financial risks.
- Portfolio Management: Chat GPT can assist portfolio managers by providing real-time information and analysis on market trends and financial data, helping them make informed investment decisions.
- Compliance and Regulation: Chat GPT can help financial institutions meet compliance and regulatory requirements by automating compliance checks and monitoring activities for potential violations.'

The software's self-portrait concludes, 'Overall, Chat GPT has the potential to greatly improve the efficiency and accuracy of various processes in the banking and finance industry, reducing costs and improving customer experience, while also helping financial institutions meet regulatory requirements and manage risk. However, it's important to note that as with any technology, **there are also potential risks and challenges, such as data privacy and security, that must be carefully considered and managed**.'

OK, thanks ChatGPT, but what wasn't mentioned there is perhaps the most winning feature when it comes to conversational design: That to train a conversational model you need huge amounts of high-quality data, which is related to the job in hand. In this case that is conversations with banking customers, and by using the billions of examples from customer data, ChatGPT saves enormous amounts of time, effort and cost in creating more realistic and effective conversational flows.

We've already heard about the increasingly cautious approach of banks, and will return shortly to those risks and challenges (perhaps not something we want to hear about in the same sentence as 'banking'). But in this whitepaper we're considering voice uses for banking, so putting aside the question of risk for a moment, how and why could this latest breed of AI and Large Language Models be such a game changer?

How voice banking works

Firstly we need to remind ourselves how the 'chain of command' works when we engage with a bank using our voice. I speak to my bank using a smart speaker, or my phone, or my glasses, or watch. What I say is very rapidly converted to a text document, and this is what the bank's system then pages through, looking for words and phrases that it 'recognizes'. That is, words or phrases that it has been trained to respond to. As we've seen from Renáta Szilágyi-Nagy's flow charting of processes, only the 'right' questions produce the 'right' answers. When this happens, the dialog coming back to the customer from the bank is also generated as a text document, which a speech engine such as Siri or Alexa then voices for us to hear. The flow is fast and usually fairly successful, within the **defined parameters** - although as we have seen with EVO Banco's voice assistant, the 'conversational effectiveness' might actually be no higher than 70%. That's good-ish, but not great, and we are all familiar with those moments when the voice app says, 'I'm sorry, I didn't understand you.' The usual level of a useful machine solution is actually reckoned to be between 85-95%, and there are numerous methods of testing the accuracy of speech models, such as the Microsoft Azure Speech Studio.

Then there is the serious issue of biases, or 'hallucinations' already mentioned, and alluded to in Google's 'cautious response' in launching Bard, ahead of the intended date. **Hallucinations are thought to be possibly even inherent to Large Language Models**, but as yet we don't know that for sure. Hallucinations are believable mistakes which may seem plausible, and can easily be skipped over, both by the system and a human (RLHF) checker. And when these mistakes are not corrected, they become part of the model, therefore further reinforcing errors. If this happens over millions of interactions, it's easy to see how things can get out of hand. Just the other day I heard from a friend who had submitted a historical article, with a little help from ChatGPT. It was well enough written, with a beginning, middle and an end, and she was about to submit it to her client when a nagging thought made her check with some more conventional fact checking sources, including Wikipedia. It turned out that the hero of the piece was not even born until after the events described - Chat GPT had hallucinated the sequence of events.

So, coming to the nub of voice banking and voice recognition, how rock solid is that in the new world of massively powerful, and readily available AI software? The tech writer Joseph Cox describes how, by using a free voice creation suite from AI voice company, ElevenLabs, he quite easily gained access to his own bank details. The ElevenLabs software allows the cloning of a voice, with intended applications such as the voicing of audiobooks and podcasts. The text to speech converter does the rest, but the ethical hacking that Cox succeeded in doing points to huge potential vulnerabilities in the 'my voice is my password' approach. We may be entering a period where fraud through deepfakes is on the rise, as the banks and the AI developers scramble to plug the holes.

But despite these possible flaws, are Chat GPT, and GPT-4 game changers? Yes, massively so, because they open the use of Large Language Models to banking and financial operators, meaning that potentially **we can all talk more easily, naturally, and increasingly *in depth* with voice bots**.

Outroduction

We hope you've found this whitepaper useful, and stimulating. The topic of voice for banking and Fintechs is a vast subject which is far from mature, with dramatic new developments in AI coming onstream in 2023, at different stages in different territories. **It's also a discipline which is changing very rapidly as AI, Machine Learning, and Deep Learning move forwards at pace.** It's reasonable to assume that the examples given here represent only a fraction of the activity that's really going on, with the probability that virtually every bank on the planet is looking at voice technologies with a view to developing and adopting something in the near future. Of course banks are not prone to announcing upcoming products and services until every detail is in place, and even those that have a voice assistant available right now tend not to major on its abilities. There are also issues of security which are now being broached as a result of the new Large Language Models coming onstream, with the possibility of voice cloning, and the potential danger of 'hallucinations'.

The future belongs to voice-based interfaces

It's a complex arena, where real expertise is needed to guide banks through the process. It goes without saying that an individual bank will only ever do a major voice assistant implementation *once*, so it's clearly a good idea to partner with agencies that have gained experience in working with banks, and specifically on voice applications. Ergomania Product Design specializes in UX-UI for the Fintech and Banking sectors. Our company is headquartered in Amsterdam and Budapest with a team comprising 50 UX and UI experts, and is active across Europe and the USA. Major achievements include the introduction of a Product Design framework for the OTP Banking Group. Among other high profile clients are BNP Paribas Fortis, KBC, TreasurUp by Rabobank, Erste, Unicredit, and Western Union.

Ergomania promotes partnership with clients, believing that **only through mutual working can satisfying solutions be created**. The company offers UX research and design, UX concepts, Service design, UI design, and Frontend development.

Ergomania's Fintech specialization spans some 10+ years with the design of websites, online administration interfaces, and internal management systems for many banks, insurers, and financial service providers. **There is a long track record of success in running distance-projects and nearshoring services** which predates the Covid-19 pandemic by many years.

Recent projects also include a number of telecommunications, energy and listing sites with provided services ranging from UX design through UI design, and on to development and delivery.

Voice User Interface, and all related voice services, are a vital part of Ergomania's offering, with high expertise and considerable experience in this area since 2012.

Ergomania is confident in stating the firm belief that the future belongs to voice-based interfaces. Traditional ways of working have changed forever, and voice-based services form an essential and growing part of every bank and Fintech business.

Are you ready for the future in voice banking?

Ergomania is.

+36 70 280 3513
hello@ergomania.eu
ergomania.eu