



THE FUTURE OF VOICE

CHERYL PLATZ

Principal Designer and Owner, IDEAPLATZ

COMPUTER, WHO IS CHERYL?

I've been designing for voice and multimodal interfaces since 2006.

@AMAZON: First designer on Echo Look and Alexa Notifications

@MICROSOFT: Multimodal design for Windows Automotive and Cortana. Dynamics 365 AI for Customer Service.

@EA & GRIPTONITE GAMES: Nintendo DS launch title (Urbz); Disney Friends DS



**VOICE USER INTERFACES
ARE THE OLDEST NEW IDEA
WE'VE ENCOUNTERED AS
AN INDUSTRY.**



TALE AS OLD AS TIME

Humans have developed the art of conversation for thousands of years.

Speech is one of the first skills we learn, and one of the last we lose.

Voice user interfaces leverage this experience to improve lives.

**The accessibility benefits are vast, and not just limited
to those with permanent accessibility challenges.**

“

My wife passed away 4 years ago leaving me, not only a widow, but a widowed quadriplegic trying to survive on his own... Alexa has been a blessing beyond my imagination. She has given me an opportunity that I never thought would be possible.

”

AMAZON ECHO REVIEW FROM MICHAEL DAVIS, FEB 2017

DESCRIBING ECHO'S AID IN HIS LIFE AS A QUADRIPLÉGIC



**IN JANUARY 2018, ONE
IN SIX AMERICANS
OWNED A SMART
SPEAKER.**

SOURCE: NPR & Edison Research

VUI: MAINSTREAM, BUT NOT MATURE

By deconstructing today's voice user interfaces, we'll find **5 key opportunities** on the path towards our future voice experiences.

OPPORTUNITY 1

Today's voice interfaces
are inherently biased.

Limited training data and an affluent user base
excludes underrepresented groups with inaccuracy.

“

“...looking at race, I found that Caucasian speakers had by far the lowest error rate. African-American speakers and speakers with a mixed racial background had higher error rates.”

”



DR. RACHAEL TATMAN (@RCTATMAN), LINGUISTICS, UNIV OF WASHINGTON
ON ACCURACY OF SIRI FOR VARIOUS DEMOGRAPHIC GROUPS
KUOW, SEPTEMBER 19 2017

DECONSTRUCTING VOICE UI BIAS

GENDER

Initial data collection is usually internal, and reflects tech demographics.

ETHNICITY

Training data expands to include early adopters, often affluent.

This may exclude underrepresented ethnicities due to wage gaps.

ACCENT

The North American focus of most of today's products mean we have yet to attain critical mass of training data for second-language speakers.

**Biased
Training Data**

BIAS SPIRAL

**High Attrition
by Excluded
Groups**

**Poor Accuracy
for Excluded
Groups**



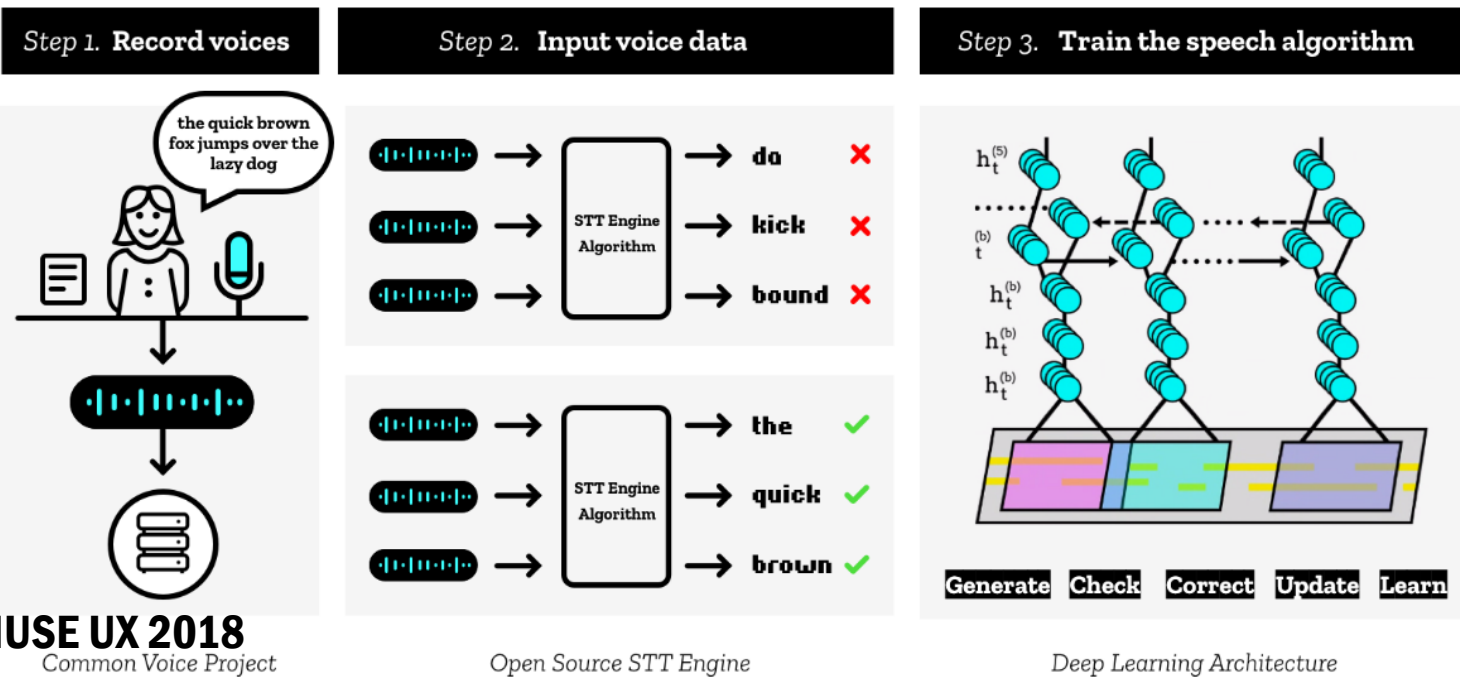
**WE MUST FIND A WAY TO BREAK THE BIAS SPIRAL,
AND MAKE THE FUTURE OF VOICE UI VIABLE FOR ALL.**

Project Common Voice

Project Common Voice by Mozilla is a campaign asking people to donate recordings of their voices to an open repository. Mozilla will release audio files and transcripts along with limited demographic information about the speakers. With a large enough data set, it's possible to train speech-to-text (STT) systems so they meet production-quality standards. The Common Voice project begins this summer, and we expect to launch the repository in the fall.

Participate in Mozilla's [Common Voice Project](#)

How a Speech Application Learns



Open source speech science is coming, and you can help.

Project Common Voice:
voice.mozilla.org

OPPORTUNITY 2

Today's voice interfaces are
reinventing the wheel.

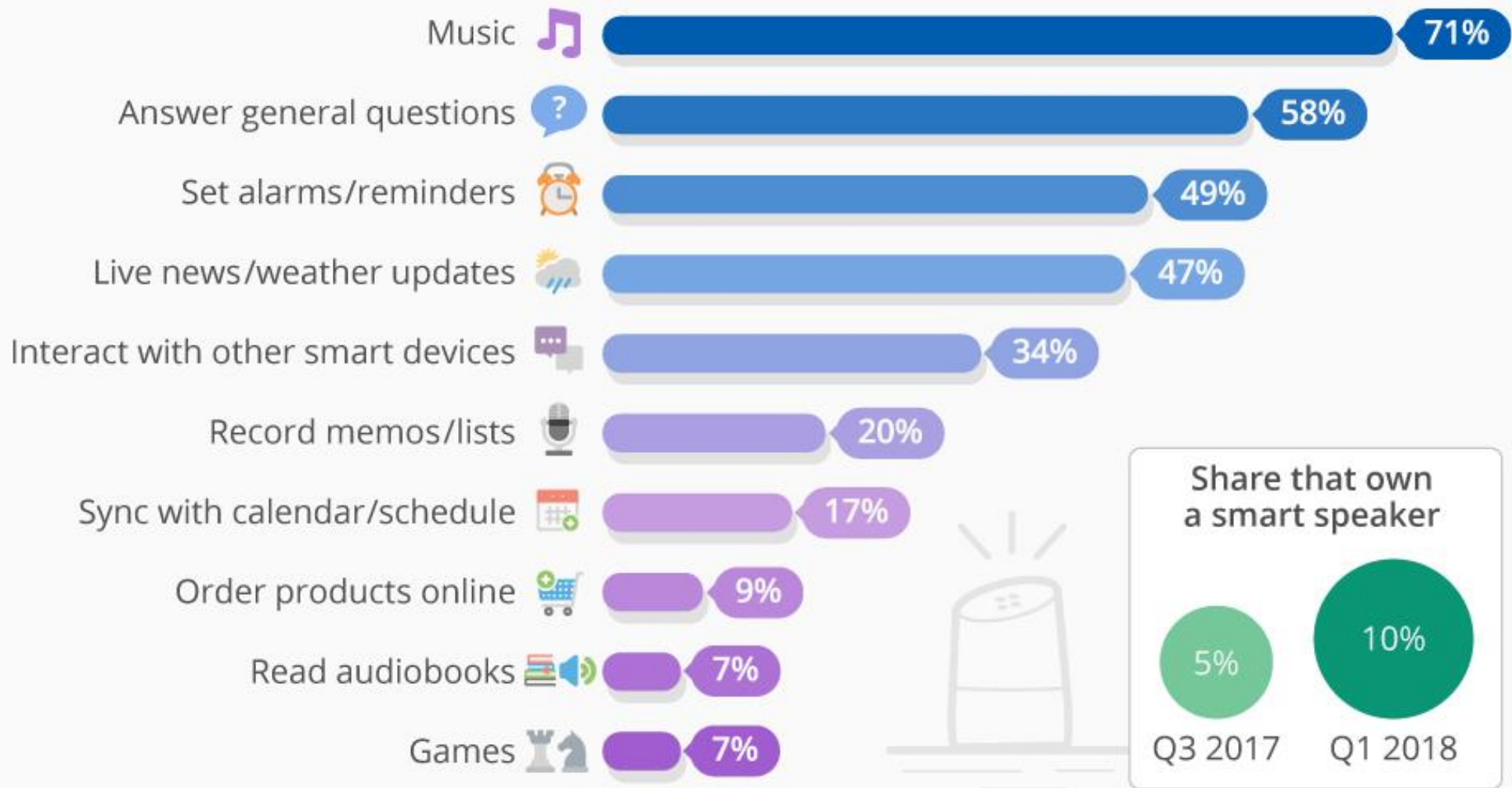
We are wasting resources re-implementing the same
basic tasks on multiple systems.

**We have an ecosystem of
voice assistants solving the
same basic problems.**



How smart speakers are used in the UK

"Which of the following do you use your smart speaker for?"





Clip from Adobe vision video: "What if you had an intelligent agent for voice editing?"

LET'S WORK TOWARDS STANDARDS

Needless differentiation of common tasks may confuse and frustrate our customers.



AMUSE UX 2018

CHERYL PLATZ // @MUPPETAPHRODITE



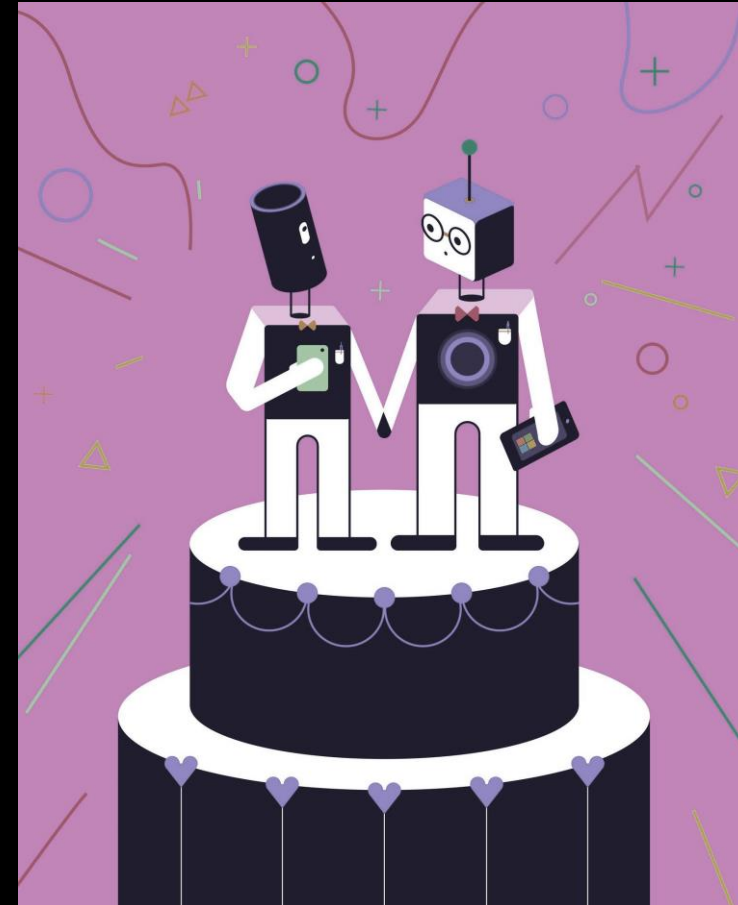
DO WE **NEED** ONE ASSISTANT TO RULE THEM ALL?

“ Through its collaboration with Microsoft, Amazon said, Alexa users will get answers to some of the same questions that Cortana can now answer – for instance, when is the next budget review with the boss?

NICK WINGFIELD, NEW YORK TIMES

AUGUST 30, 2017

ILLUSTRATION: MENGXIN LI





**LET'S BUILD A CHOIR OF HARMONIOUS
VOICE INTERFACES TOGETHER.**

AMUSE UX 2018

CHERYL PLATZ // @MUPPETAPHRODITE

OPPORTUNITY 3

**Most voice UIs are
barely conversational.**

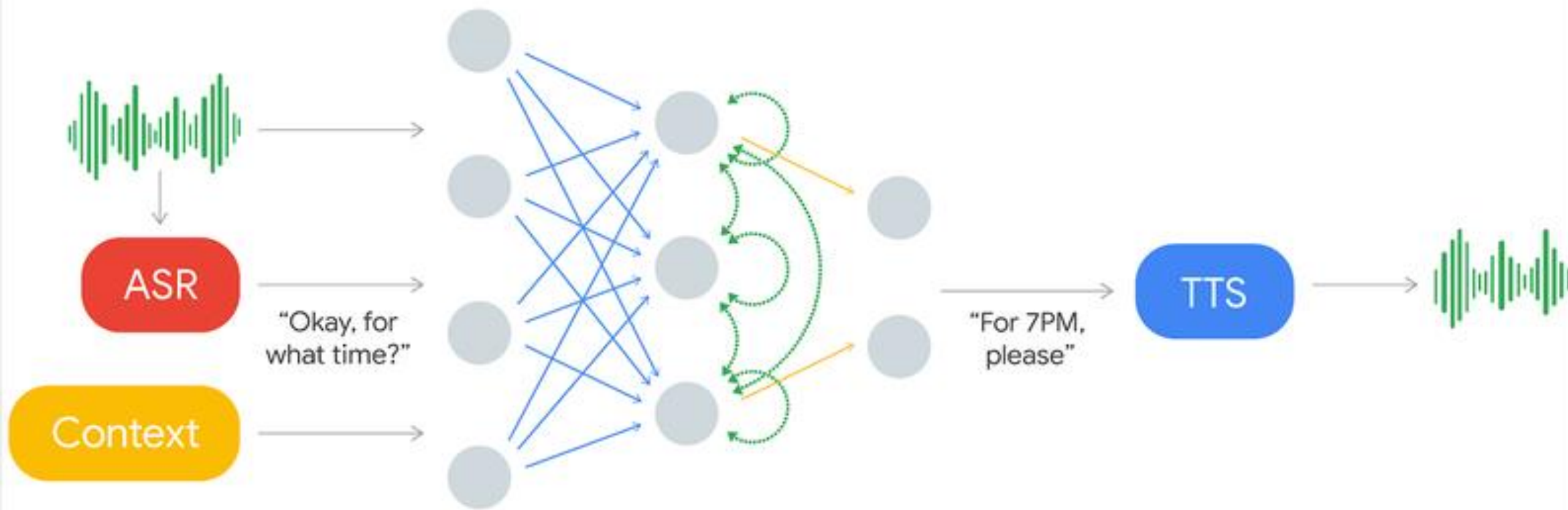
**Our brains expect these voices to converse in human
ways - but most assistants are still static, not adaptive.**



**IT LOOKS LIKE YOU MIGHT BE IN THE
AWKWARD EARLY STAGES OF
CONVERSATIONAL UI. CAN I HELP?**

PLEASE NO

RUN AWAY



**THERE'S BEEN RECENT PROGRESS, LIKE DUPLEX...
...BUT CONVERSATION IS MORE THAN A TRANSACTION.**

“

...The psychology of interface speech is the psychology of human speech: voice interfaces are intrinsically social interfaces. Designers must create voice interfaces for brains that are obsessed with extracting as much social information as possible from speech.

”

CLIFFORD NASS AND SCOTT BRAVE, “WIRED FOR SPEECH”, 2005



Clip from "Her": Warner Brothers / Anapurna Pictures



RESEARCH SHOWS HUMANS DON'T DISTINGUISH DIGITAL SPEECH FROM HUMAN SPEECH, WHICH BRINGS TREMENDOUS CONSEQUENCES.

WIRED FOR SPEECH: THE TIP OF THE ICEBERG

CONSISTENCY & GENDER

Participants responded far more positively when gendered voices behaved in ways that aligned with their society's gender stereotypes.

IE, a male voice narrating a description of power tools.

SIMILARITY

Participants were more responsive and more trusting when presented with a voice UI that **matched their own personality**, particularly **introversion-extroversion**.

COMPANIONSHIP IS CALLING

“Microsoft has turned Xiaoice, which is Chinese for “little Bing,” into a friendly bot that has convinced some of its users that the bot is a friend or a human being.”

The Verge
May 22, 2018

MICROSOFT TECH ARTIFICIAL INTELLIGENCE

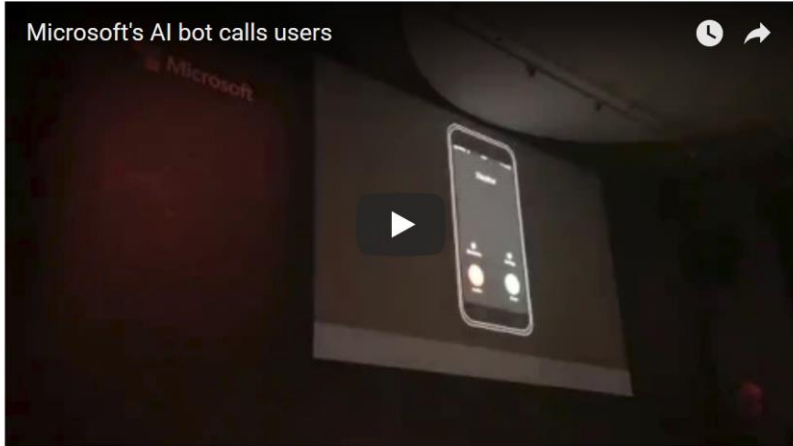
Microsoft also has an AI bot that makes phone calls to humans 50

Similar to Google Duplex, but only in China

By Tom Warren | @tomwarren | May 22, 2018, 8:43am EDT

f t SHARE

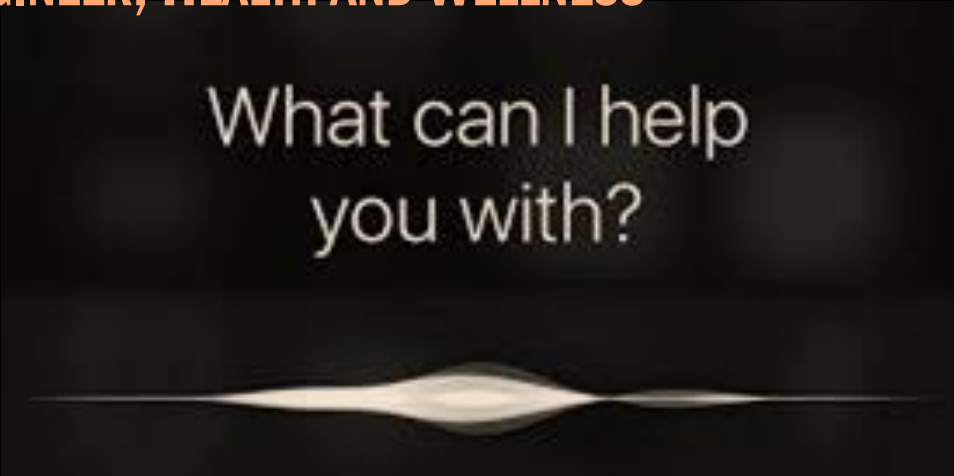
Microsoft's AI bot calls users



“ People have serious conversations with Siri. People talk to Siri about all kinds of things, including when they’re having a stressful day or have something serious on their mind. They turn to Siri in emergencies or when they want guidance on living a healthier life. ”

APPLE JOB POSTING, SIRI SOFTWARE ENGINEER, HEALTH AND WELLNESS

APRIL 4, 2017



What can I help
you with?



Clip from "Her": Warner Brothers / Anapurna Pictures

WHAT DOES A RELATIONSHIP LOOK LIKE?

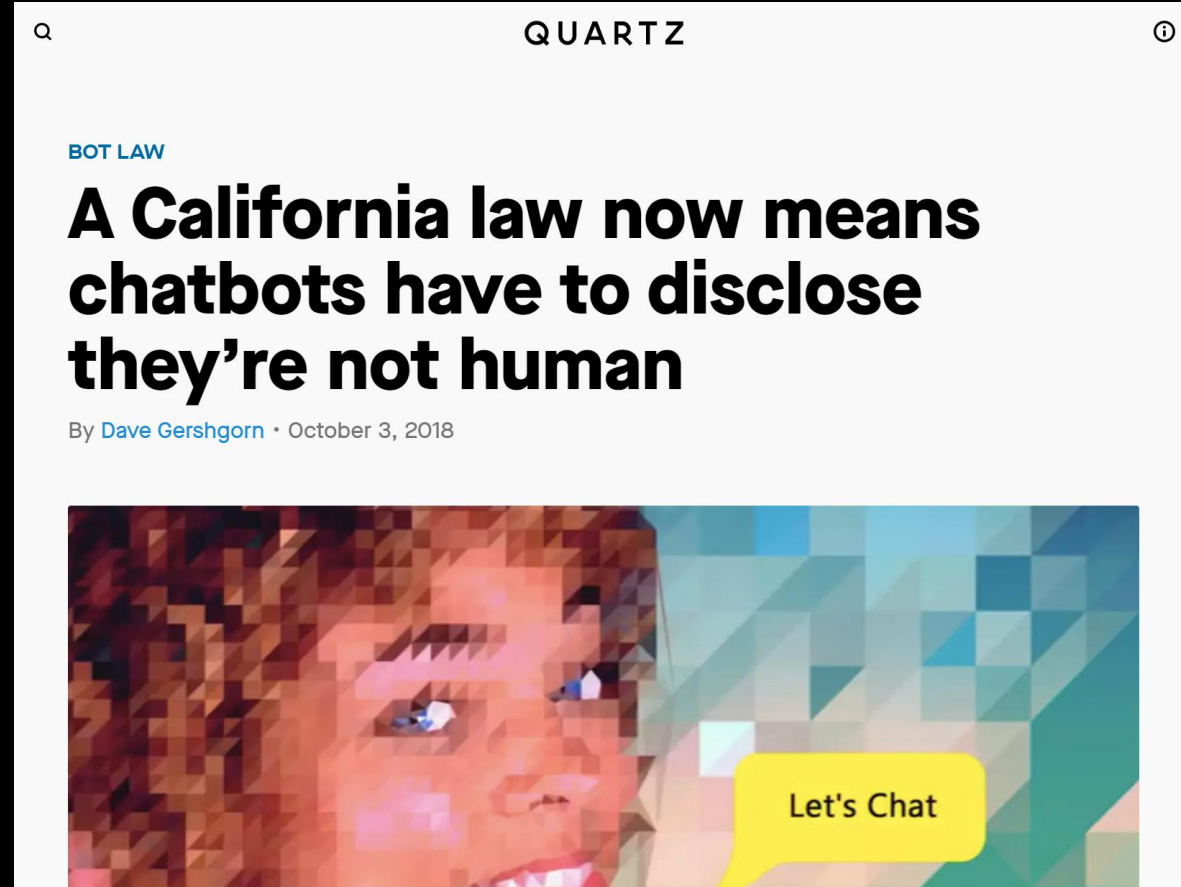
- How can we (ethically) model a relationship over time?
- What information is saved, and for how long?
- What level of transparency and control is required?
- Does the assistant's personality adapt, or remain fixed?
- How does politeness (or lack thereof) affect interactions?

REGULATING THE ILLUSION OF LIFE

“The new law goes into effect on July 1, 2019...

It will require companies to **disclose whether they are using a bot to communicate with the public on the internet** (something like ‘Hi, I’m a bot.’) ”

- Dave Gershgorn, Oct 3 2018



CONVERSATIONAL CONSENT MATTERS.

We have a responsibility to clearly inform customers when they are speaking to conversational AI, and to allow them to opt out.

Any hope of building trust with customers – and as an industry – requires we respect a customer’s right to choose.

OPPORTUNITY 4

Voice interfaces don't
yet help us with
complex work.

Creative and large-scale tasks aren't meaningfully supported.



MAJOR PRODUCTIVITY CHALLENGES

ENVIRONMENT

Not all productivity tasks occur in secure or isolated spaces, especially with the advent of open workspaces.

USER TAXONOMIES

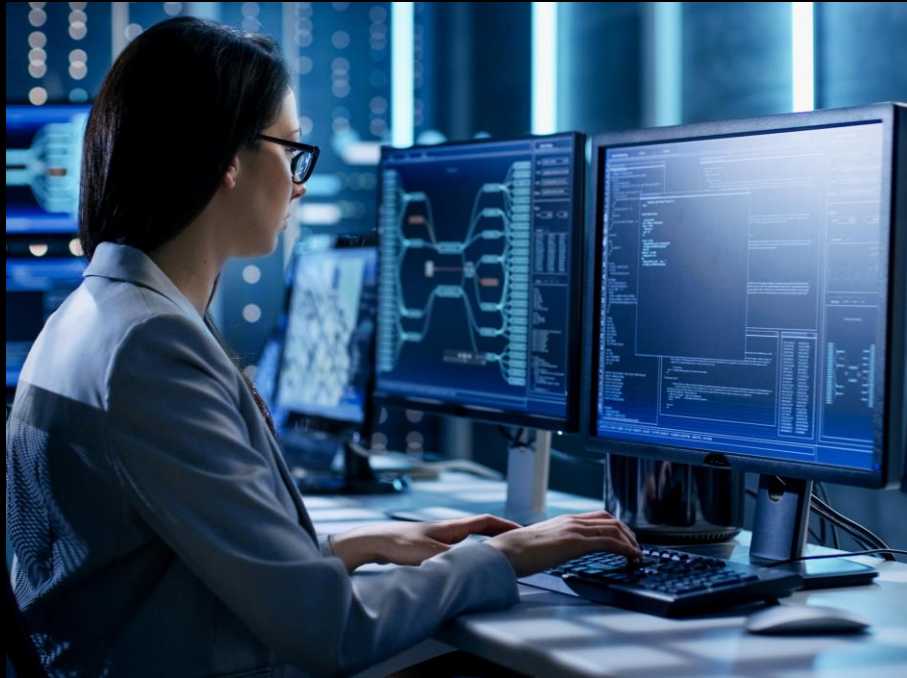
Customer-defined object names aren't guaranteed to be acoustically unique.

DESIRABILITY

We don't yet fully understand what tasks customers are willing to complete without visual confirmation.

WHERE TO START WITH VOICE PRODUCTIVITY?

Provide peace of mind via monitoring



Solve “needle in a haystack” knowledge problems



FUTURE OF VOICE PRODUCTIVITY

Contextual manipulation

Help customers identify the object they need from large data sets using **semantic identifiers**, rather than names.

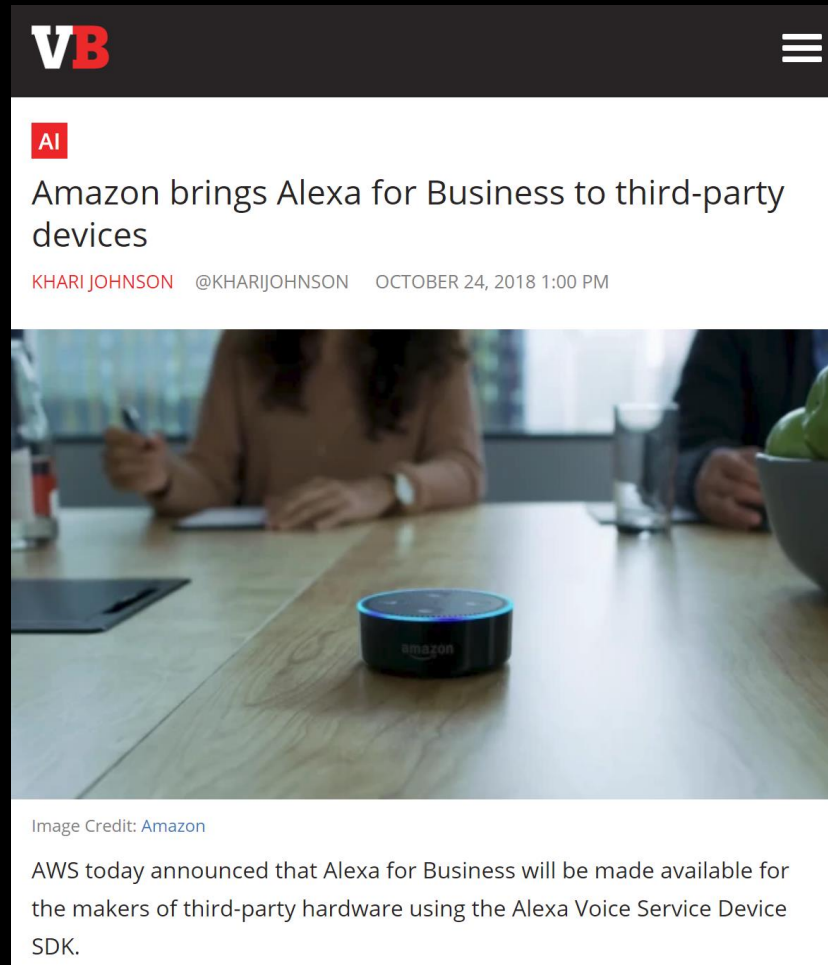
Conversational authoring

Create and edit new documents from scratch – by **describing the content**, instead of navigating UI.

VOICE IN BUSINESS IS COMING... SOON?

**Alexa for Business:
rolling out to help
with tasks like
meeting scheduling
on a variety of
devices.**

**Others will follow suit
once this becomes
the norm in offices.**



The screenshot shows a news article from 'VB' (VentureBeat) with a red 'AI' tag. The headline is 'Amazon brings Alexa for Business to third-party devices'. The author is 'KHARI JOHNSON' (@KHARIJOHNSON) and the date is 'OCTOBER 24, 2018 1:00 PM'. Below the text is a photograph of an Amazon Echo smart speaker on a wooden conference table. In the background, two people are seated at the table, one holding a pen and the other with a bowl of fruit. The article text below the image states: 'Image Credit: Amazon' and 'AWS today announced that Alexa for Business will be made available for the makers of third-party hardware using the Alexa Voice Service SDK.'

OPPORTUNITY 5

We have multiple input modalities, but few multimodal systems.

To fully realize technology's potential, we must design flexible cross-modal systems from the ground up.

MULTIMODALITY MAXIMIZES INCLUSIVENESS

Voice interfaces **do** change lives for customers who were not well served by primarily visual UI.

However, we shouldn't leave the deaf and others with auditory impairments behind in our march towards a bold new world.

MODELS OF MULTIMODAL INTERACTIVITY

SEQUENTIAL

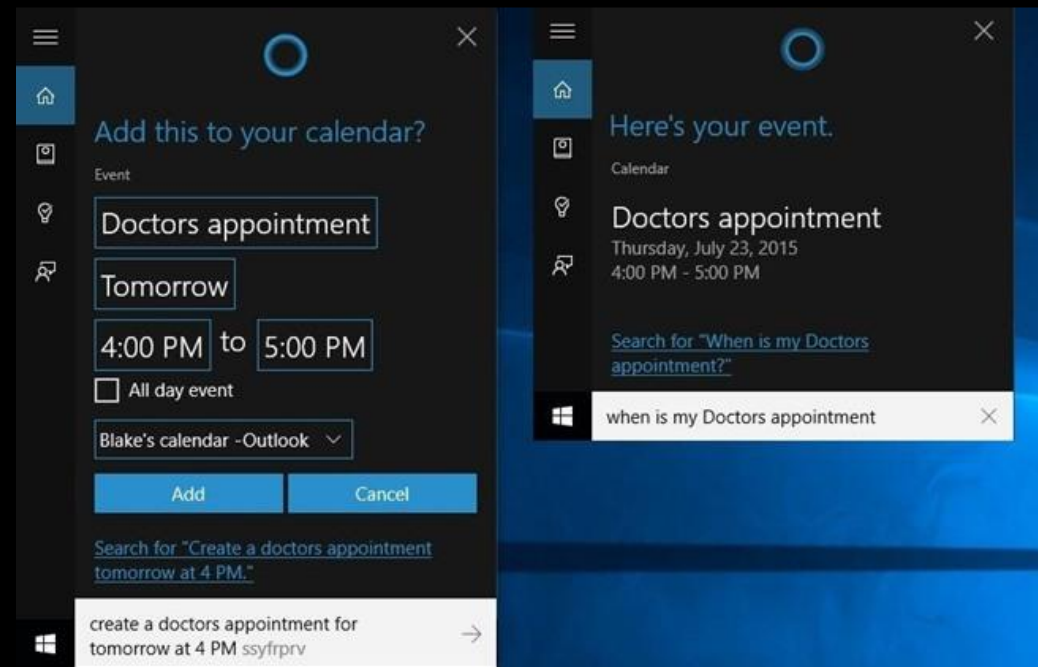
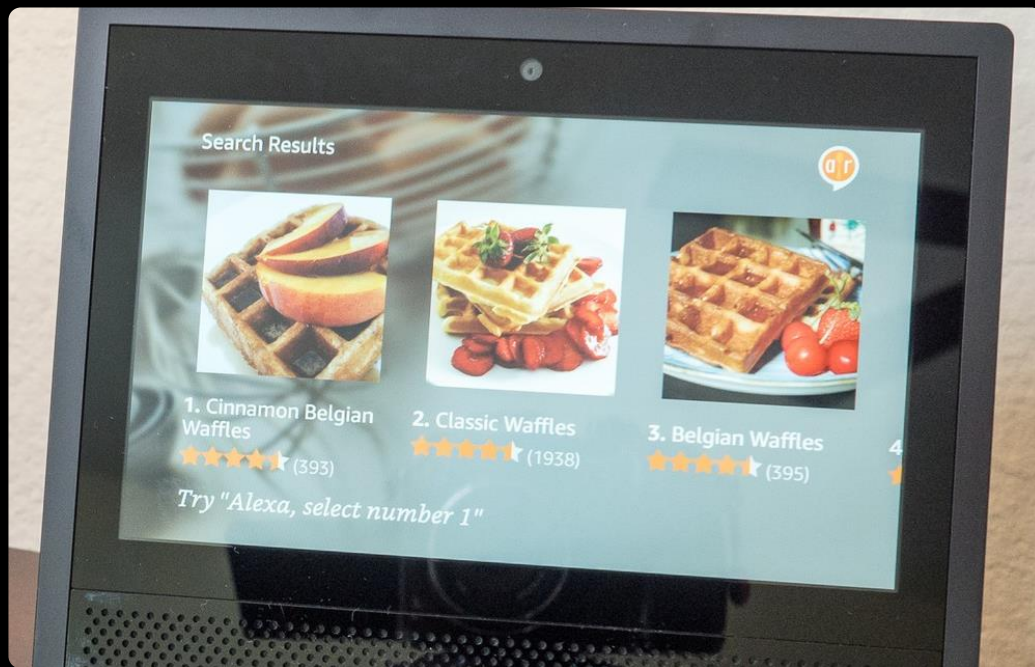
Multiple input modalities are supported, but only one at a time. Ideally, state is saved upon changing input – but not always.

SIMULTANEOUS

Support for multiple input modalities at once, which can be processed in tandem to accomplish more in real time.

Example: Pointing to a spot on a map and saying “How do I get there?”

TODAY'S SEQUENTIAL MULTIMODALITY



SIMULTANEOUSLY MULTIMODAL SYSTEMS

Centralized state

Customer state in a specific scenario is stored globally, to allow seamless switches between devices and input modalities.

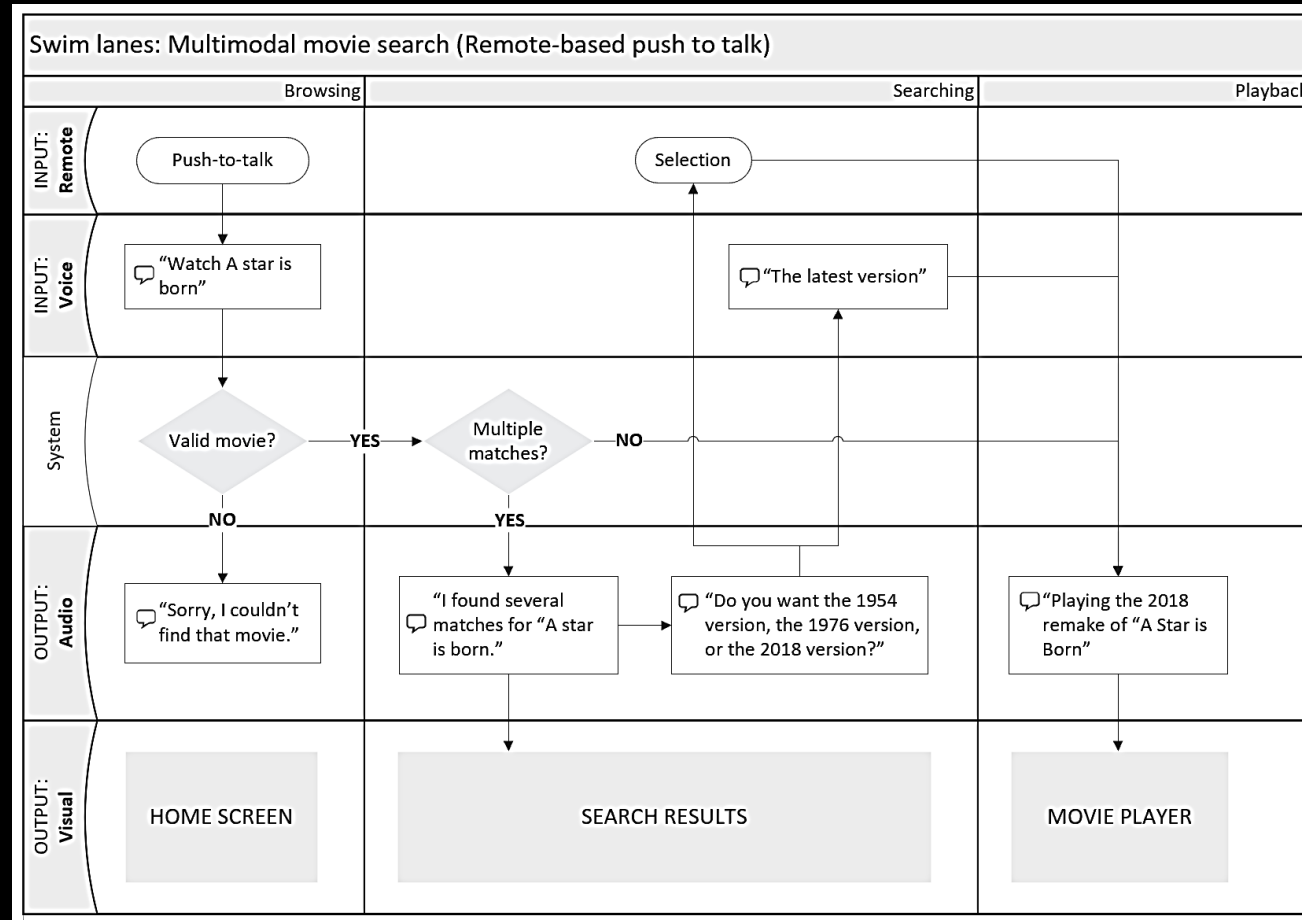
Highly contextual

Customer's most recent input mode
Time of day
Current location
Ambient noise levels

Adaptive output

Content scales to match the output medium: spoken prompts shorter than written prompts.
Guide customer to most likely input success.

HOW DO WE ADAPT OUR DESIGN PROCESS?





**IN A WORLD OF AR AND VR, SIMULTANEOUS
MULTIMODALITY WILL BECOME EVEN MORE IMPACTFUL.**



**WHAT AWAITS US IN A
FUTURE OF VOICE UI?**



ENHANCED PRODUCTIVITY

Help

14:21

Alexa play my favorite playlist

© Prelude Suite, Op. 94 - No. 3

⏮ Previous track

✓ Prelude Suite, Op. 94 - No.

⏭ Next track



MULTIMODAL ADAPTIVITY

00:42

26/53

-01:59



INCLUSION AND COMPANIONSHIP

With responsible design, voice user interfaces will unlock new opportunities and a new era in human empowerment.

Ready to get started? Join me on Wednesday for my full-day workshop, “Giving Voice to your Voice Designs.”

**AS WE PUSH TOWARDS THIS
SCIENCE FICTION VISION OF THE
FUTURE...**



LET'S BUILD A FUTURE OF
VOICE INTERFACES WHERE
OUR HUMANITY IS **AMPLIFIED**,
NOT **ATROPHIED**.



May the voice be with you.

Workshops & more at ideaplatz.com

CHERYL PLATZ

Principal Designer and Owner, IDEAPLATZ

Twitter & Medium: @MuppetAphrodite