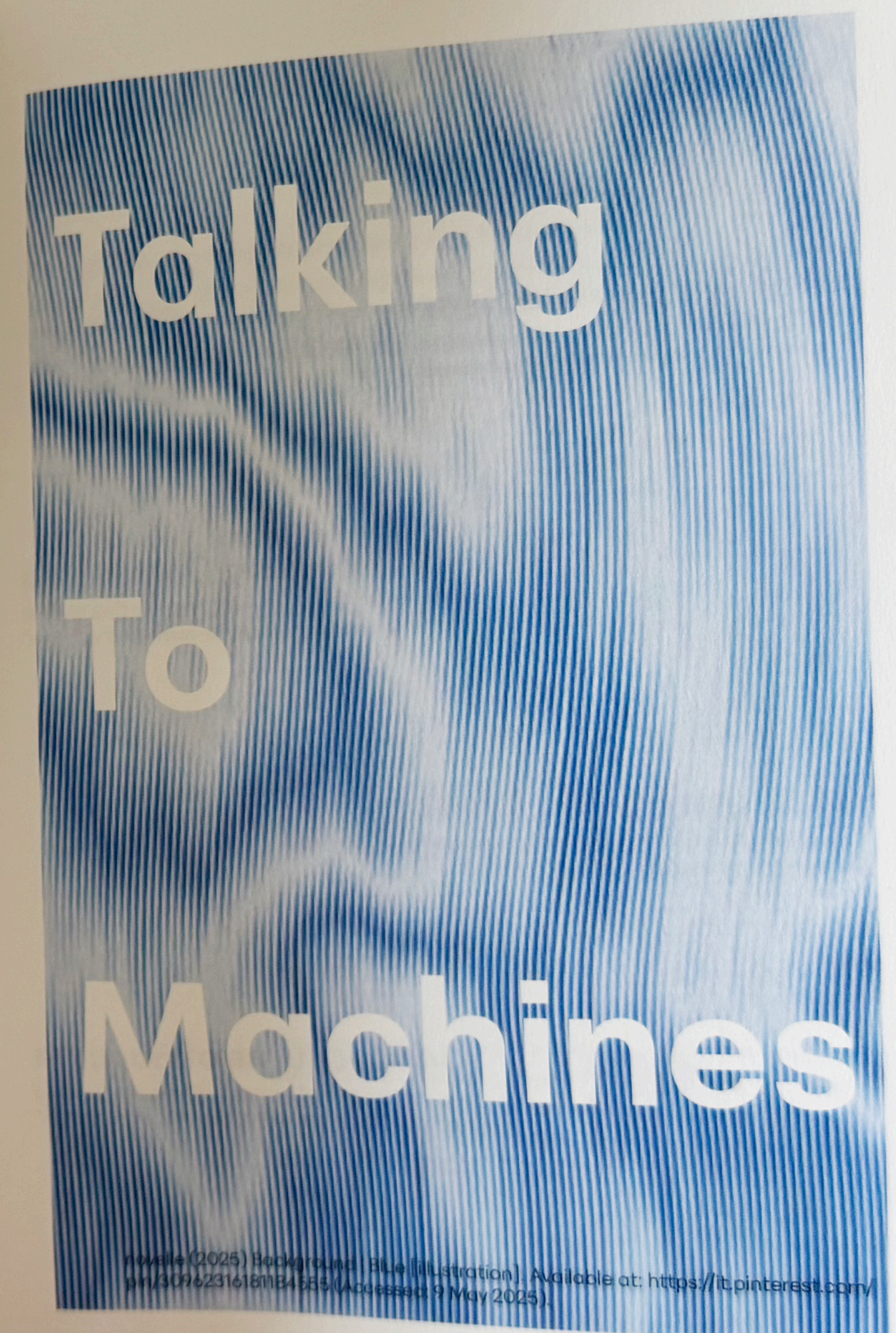
Assistant for human #1

Talking To Machines

Tom Geller

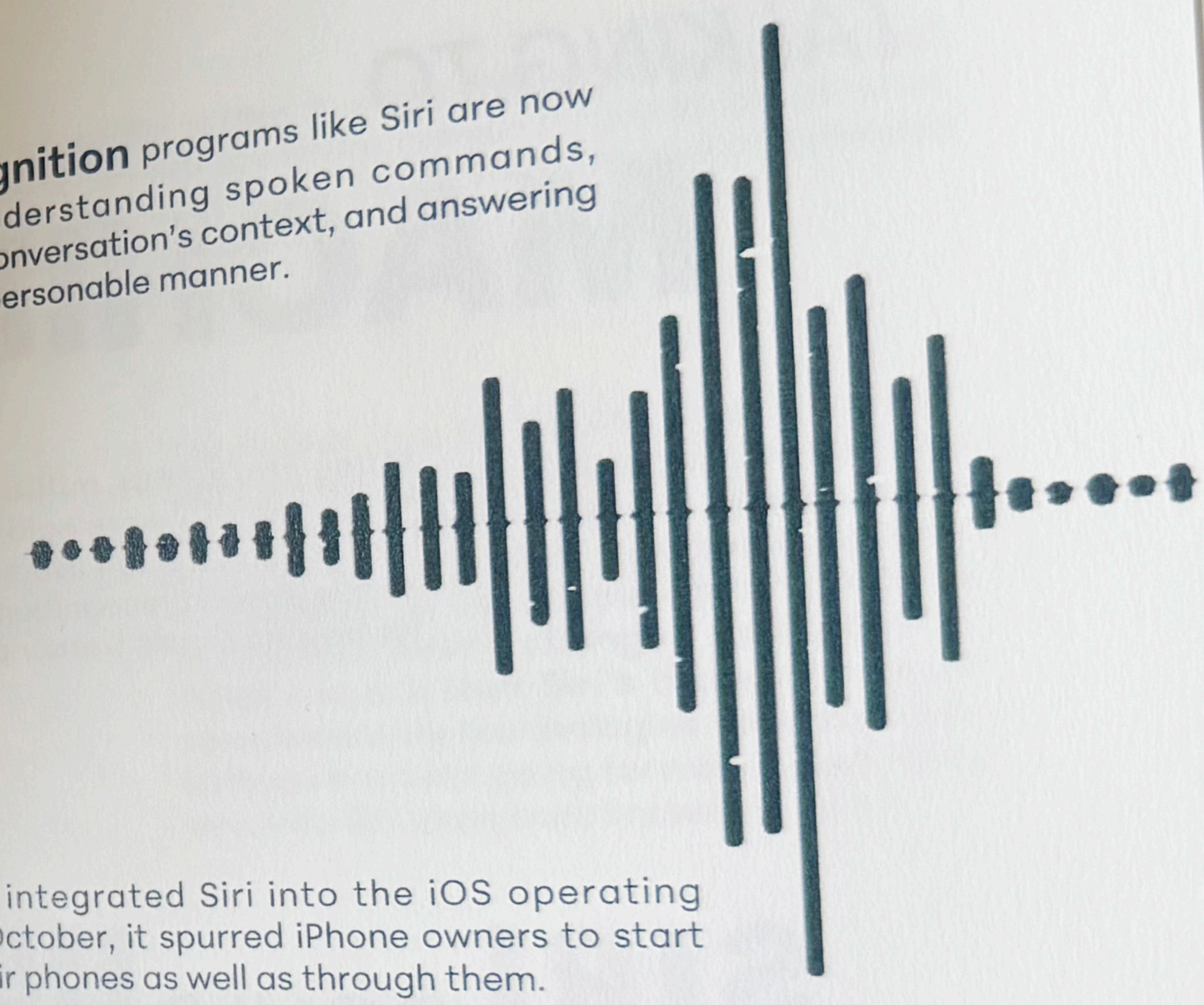


[illegible] (2025) Background | Blue [Illustration]. Available at: https://it.pinterest.com/pin/[illegible] (Accessed 9 May 2025).

CONTENTS

Talking to Machines

Voice recognition programs like Siri are now capable of understanding spoken commands, recognizing a conversation's context, and answering questions in a personable manner.



When Apple integrated Siri into the iOS operating system last October, it spurred iPhone owners to start talking to their phones as well as through them.

The program, which converts spoken commands such as "Schedule dinner with Lisa at 6 tonight" into calendar appointments, Web searches, and the like, is the most widely distributed example of a cognitive assistant to date.

TALKING TO MACHINES

More than four million iPhone 4S's featuring Siri were sold during its first weekend. Although users might see it as simple speech recognition, its abilities go far beyond simple transcription.

SIRI COULD HELP

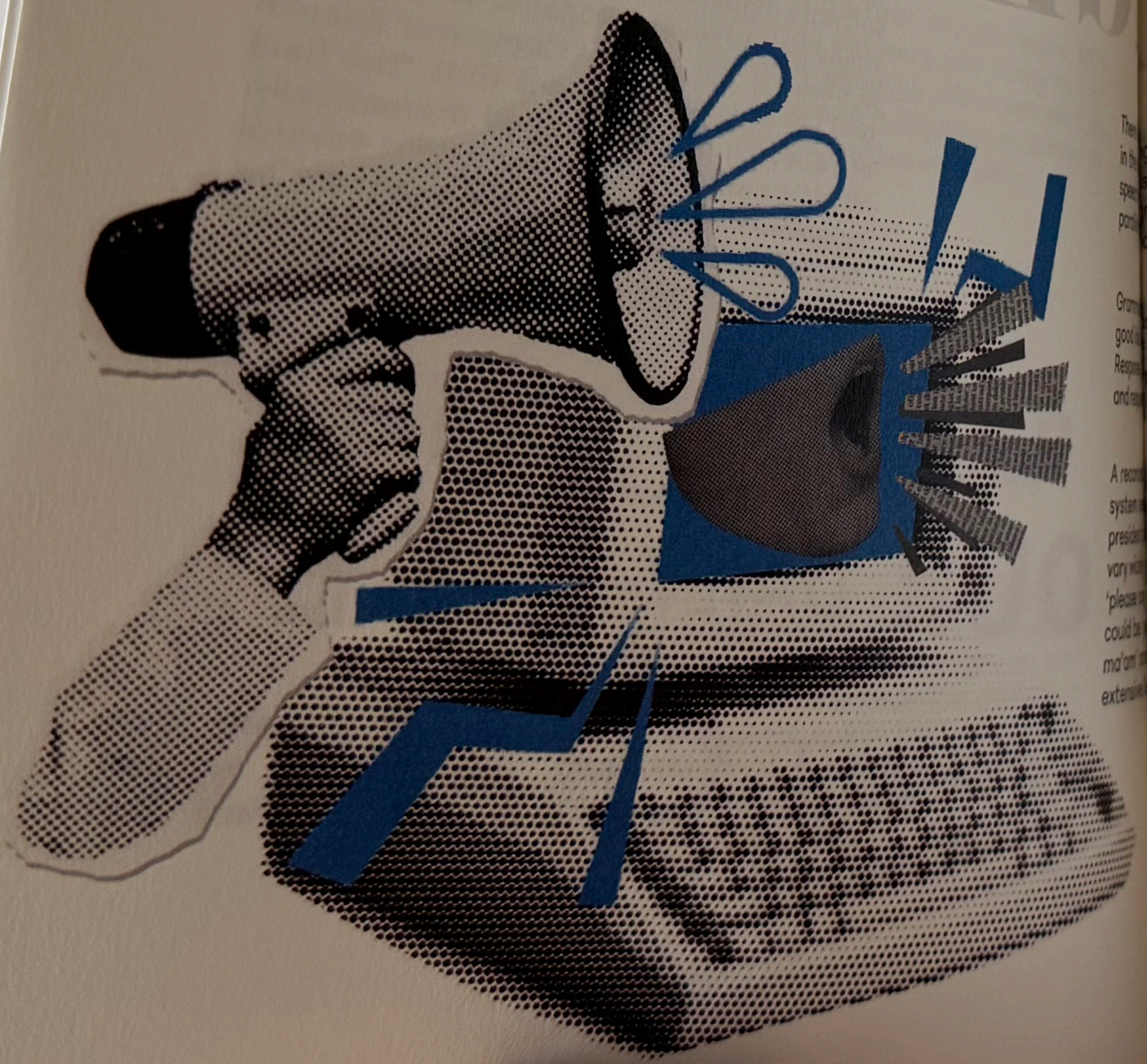
Siri represents an important moment when voice recognition, information management, artificial intelligence, task fulfillment, and user interface marry in a way the general public finds usable and productive. As Wolfram Research executive director Luc Barthelet says, "

The news about Siri is that it works. People have tried to get computers to answer questions conversationally for at least 15 years, but only now has the technology reached a threshold where people overall like it.

The iPhone's popularity also gives intelligent software assistants wider exposure than they would get otherwise. Roger K. Moore, editor in chief of the journal Computer Speech and Language, points out that "the field of research hasn't changed dramatically. What's new is that Siri's brought several complementary technologies together. Our business has been going for many years. Only now, with Siri, everybody knows about it."

OUR DAILY LIFE COULD BE CHANGE

1 Executing Commands



The… road between the spoken command and its fulfillment, though. The first step in th… is to convert the audio of speech into meaning. The two main applications of spee…nition—dictation and command recognition—have forced researchers to pursue par…ods that balance vocabulary, accent, and context needs.

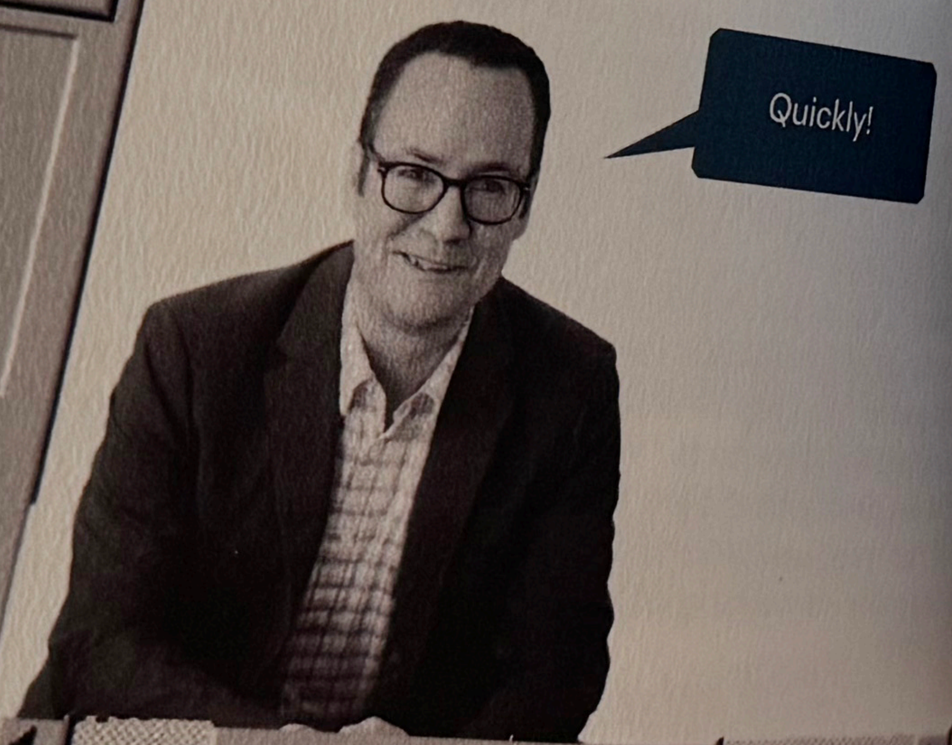
Gram…sed voice recognition is optimized for situations where the program has a very good… what the speaker will say. Its most common application is in Interactive Voice Respo…) systems, such as those that some airlines use to interpret spoken reservations and re… for information. These are often conversational.

A reco…oice asks the speaker a question, then listens for the response. As a result, the system… to understand only a limited vocabulary. But according to Dan Faulkner, vice presid… product and strategy for the Enterprise Business Unit at Nuance, responses can vary w… matter how restricted the domain is. "When a phone prompt system tells you to 'please…es or no,' we might be pretty confident the speaker will say 'yes' or 'no.' But 'yes' could b…h,' 'that's correct,' or 'yup,' and in the Southern states some people will say 'yes, ma'am' …o, ma'am.' So even for something as simple as 'yes' or 'no,' you need quite an extensive… of phrases."

Contrast

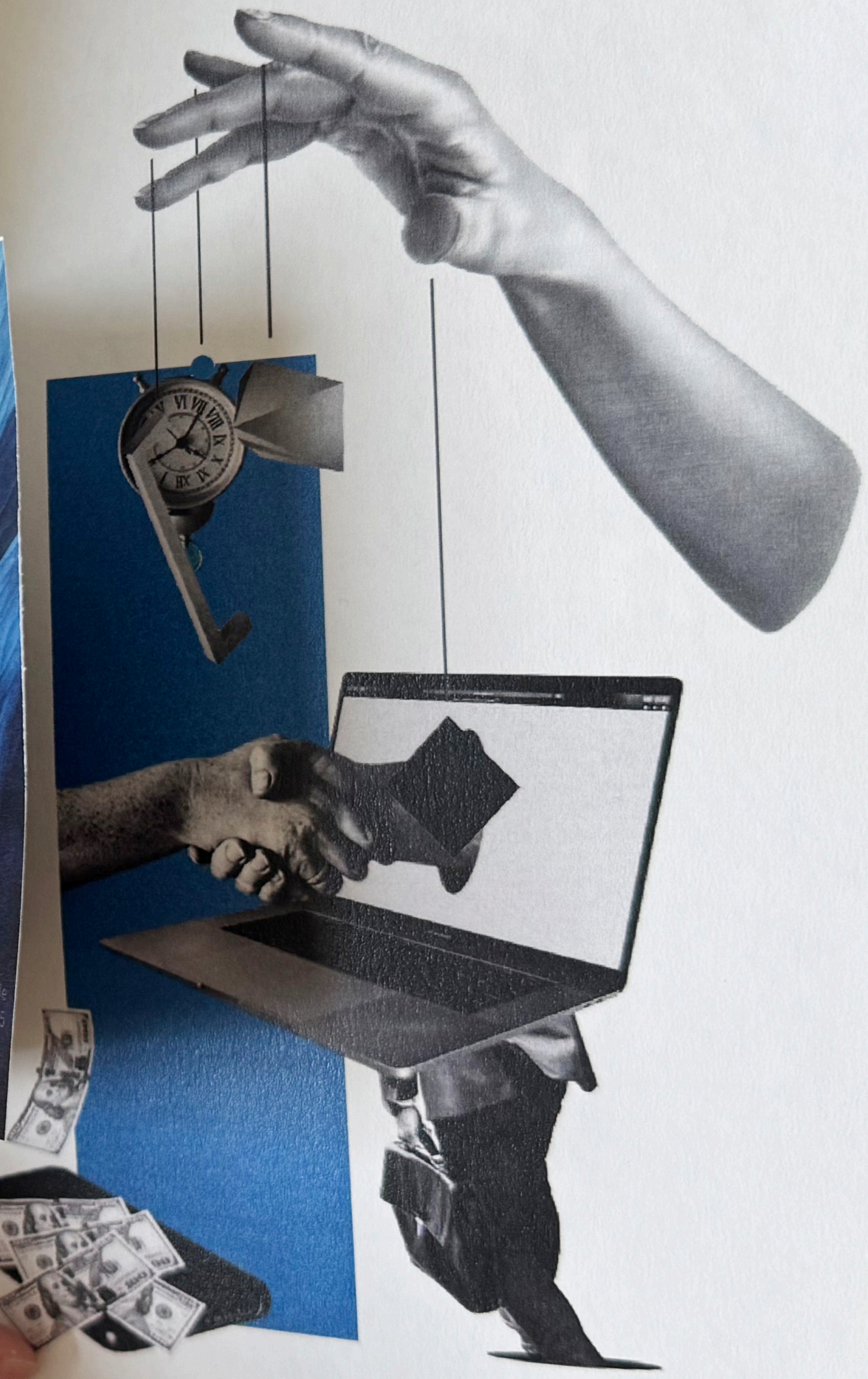
A language-based voice recognition system is optimized for dictation. It makes few assumptions about context, attempting to recognize and transcribe every word it hears. In gaining a wide vocabulary, it gives up the ability to understand a variety of accents.

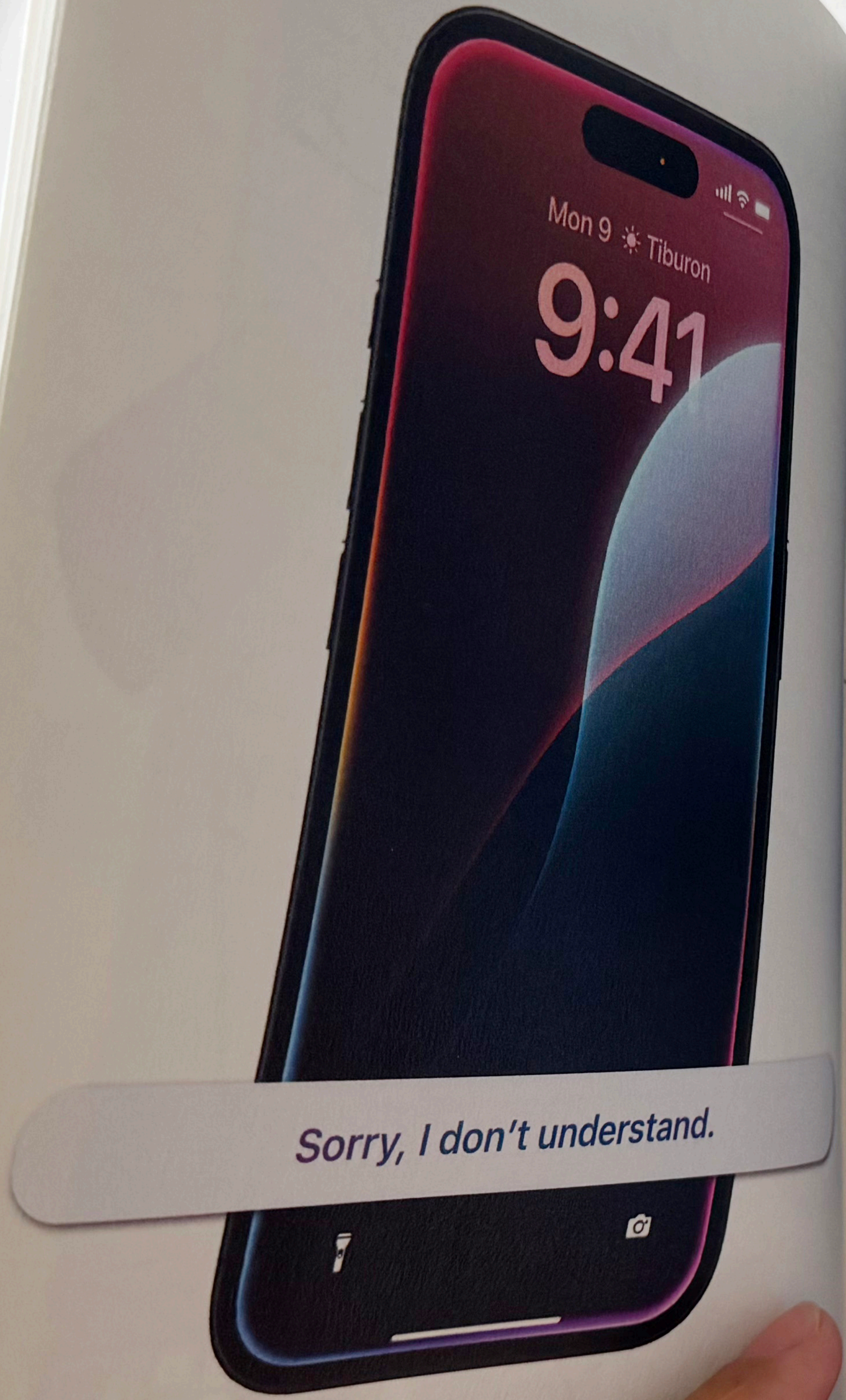
According to Peter Mahoney, general manager of the division of Nuance that produces its Dragon Naturally Speaking products, modern dictation programs no longer need the 30–60 minute initial training period that earlier versions did.



For Help, click ... Menu.

Black believes Siri delivers that satisfaction partly through its helpful-yet-sassy tone. "It doesn't just answer questions," he says. "It has a character. It wants to name you, to know who you are.



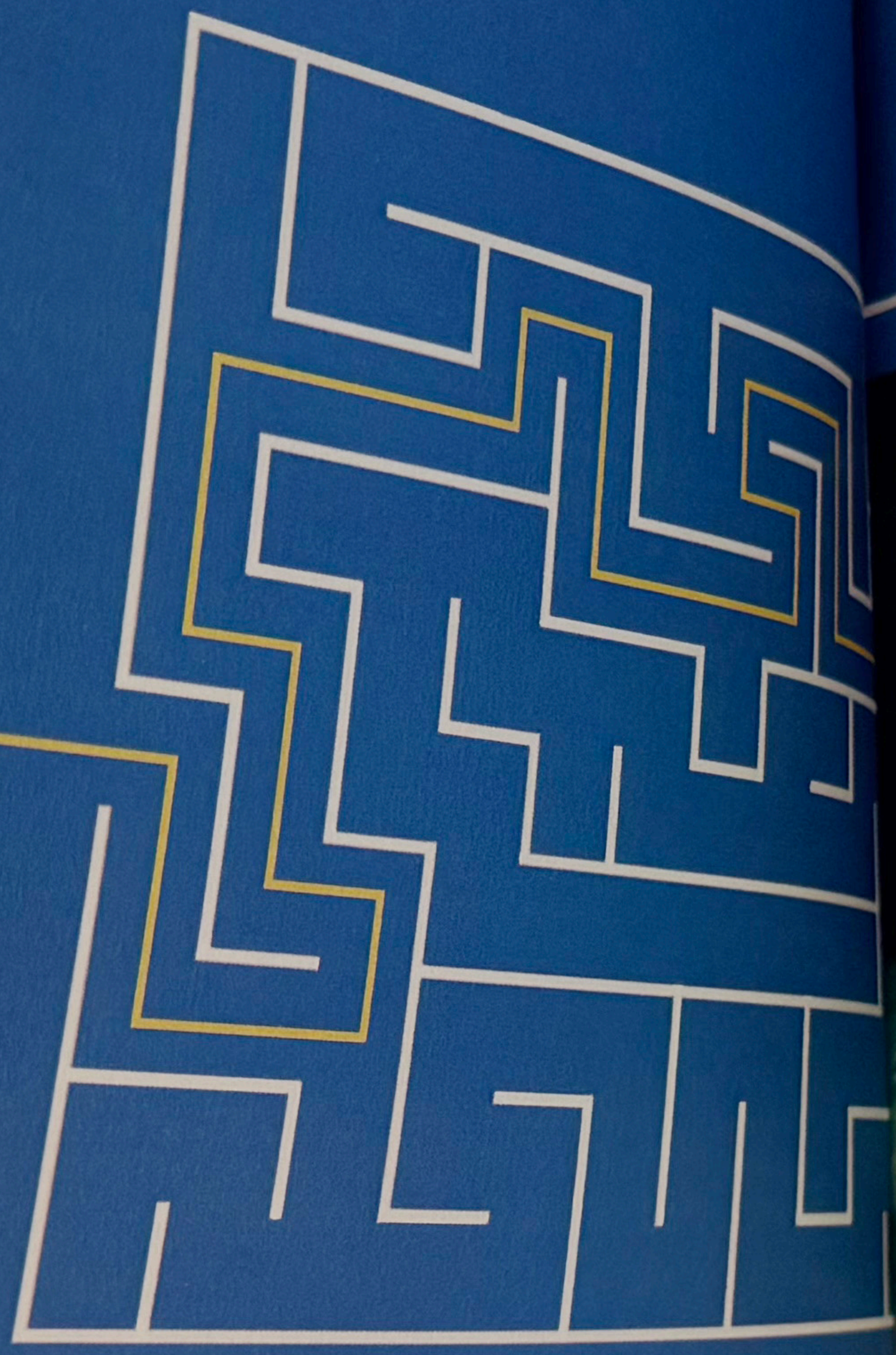


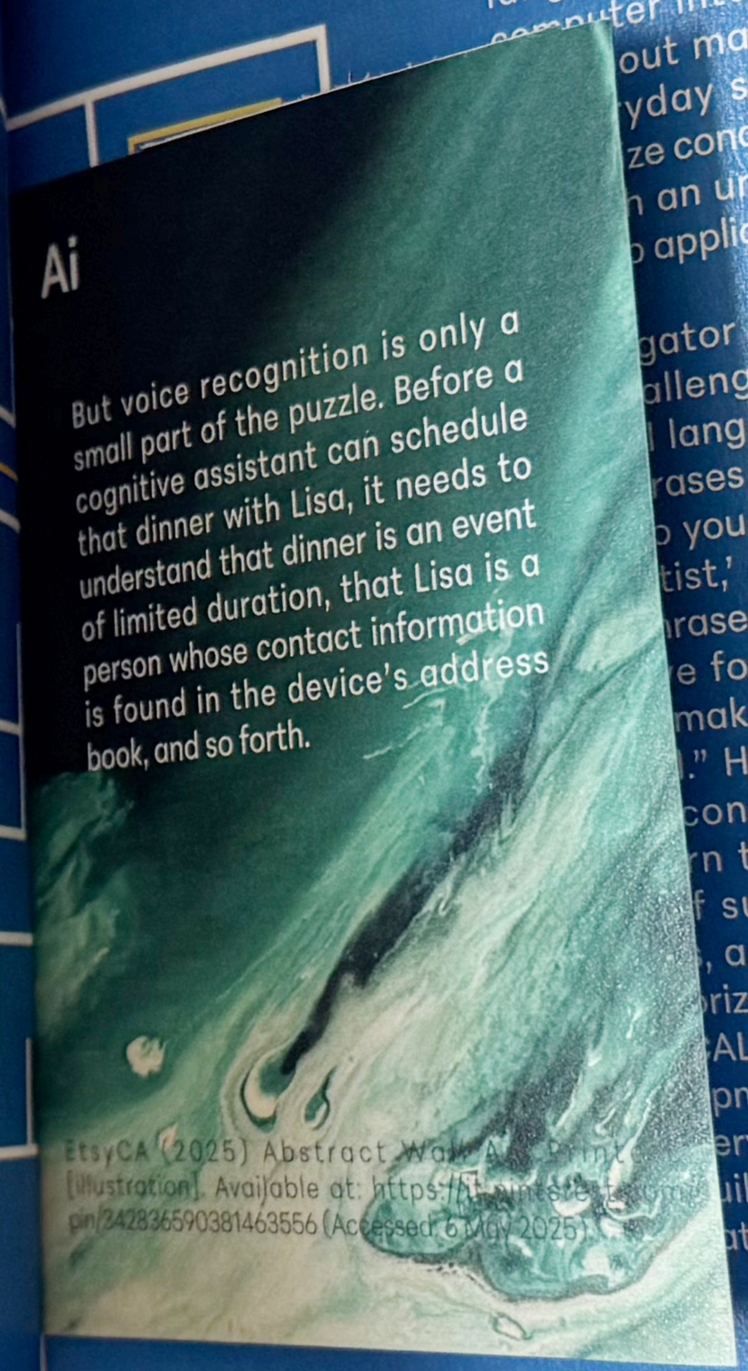
Learning and Organizing

[illegible] [illustration]. Available at: [illegible]/2TG9Dwanf (Access[illegible]

Puzzle

But voice recognition is only a small part of the puzzle. Before a cognitive assistant can schedule that dinner with Lisa, it needs to understand that dinner is an event of limited duration, that Lisa is a person whose contact information is found in the device's address book, and so forth.





Some of this understanding came from a Defense Advanced Research Projects Agency (DARPA)-funded project, Cognitive Assistant that Learns and Organizes (CALO), which was part of DARPA's PAL (Personalized Assistant that Learns) program. CALO's focus was not on voice recognition per se, or natural-language understanding, or on human-computer interaction in general. Rather, it … out making computer systems learn … yday settings, such as learning to … ze concepts. It then had to relate them … n an underlying ontology, and trigger … applications and Web services.

… gator C. Raymond Perrault describes … allenge of solving for ambiguities in … language. "Movie titles are typically … rases in the language," Perrault points … you could say 'Get me two tickets to … tist,' and the system would recognize … rase as a movie title. On CALO we tried … e for such ambiguities robustly, and … make it easy to build systems that do … ." Humans resolve these ambiguities … context, while teaching a computer … rn them requires building very long … f such things as addresses, people, … , and organizations, along with a solid … rization system to manage them. Ideas … ALO suggested a new approach to the … pment of a spoken interface to a set of … ervices, eventually realized in Siri, which … uilt by a company that spun out of SRI … ational.

Puzzle

Neither a robust ontology nor highspeed voice re tion are possible without substantial data as input. Here, the Web's social nature—and some caref ded clauses in enduser license agreements—allow software assistants to collect acoustic, syntactic, a ctual information. **"It wouldn't be possible to do something like Siri 10 to 15 years ago because y uldn't get enough data to train the system,"** says Alan W. Black, associate professor at the Languag chnologies Institute of Carnegie Mellon University. "Google started collecting data years ago throug ee 411-GOOG informational service. Notably, they advertised on billboards rather than online. What were actually doing was finding out how ordinary people asked questions."

By contrast, Nuance's Faulkner recalls how the comp y trained its dictation products to understand noisy phone transmissions in past years. "We'd pay peop o come into the office and give them scripts. We'd give them a mobile phone, put them in a cab, tell th to call a number, then record their speech."

Oc… no name [illustration]. Available at: https://it.pinterest.com/
pi… Accessed: 9 May 2025).

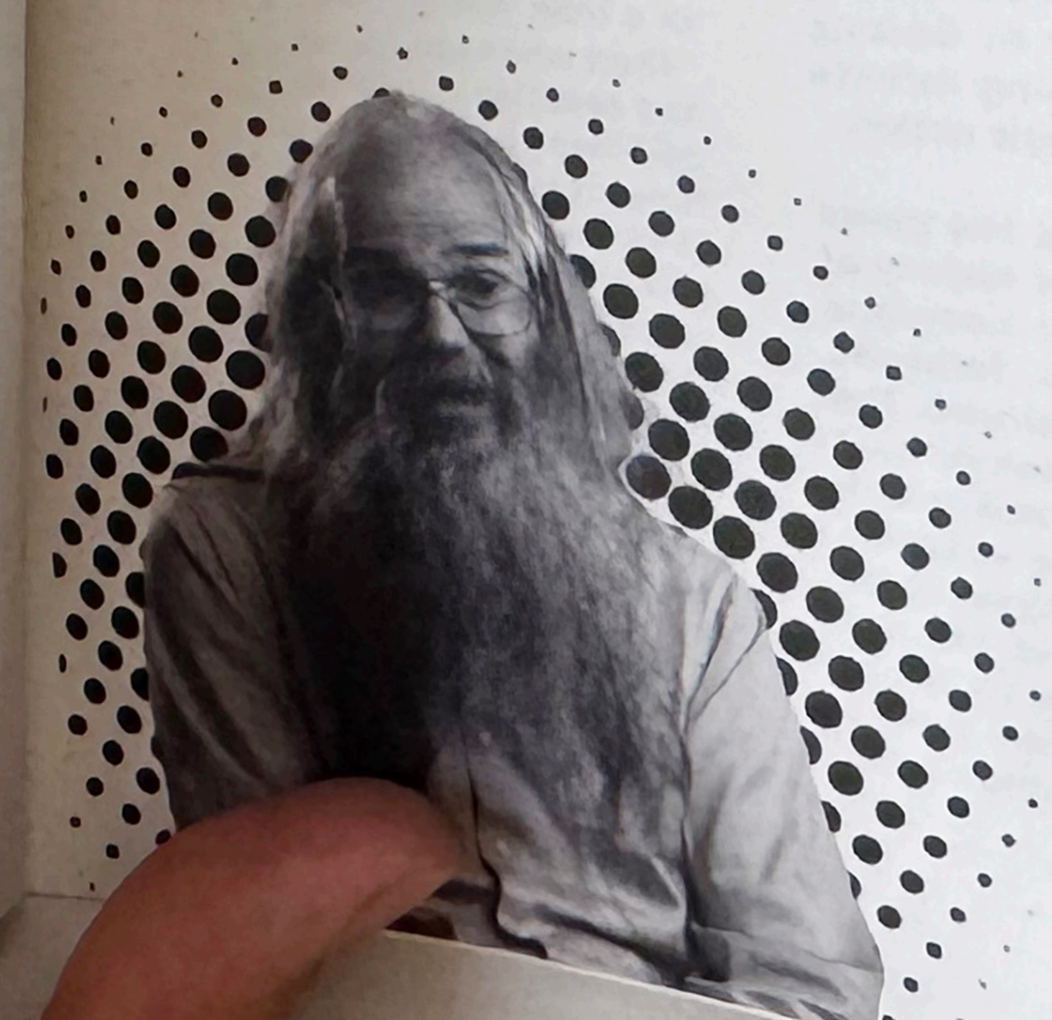
Measures of Success

Teak New York (2025)
(Accessed: 11 April 202

Measures of Success

Ultimately, these steps serve the single goal of delivering a relevant, true, and useful response with acceptable speed. But as Carnegie Mellon's Black points out, a software agent's job is not done if it only delivers facts. "One standard measure of spoken dialogue systems is task completion," he notes. "Did the user successfully get the weather? But it's clear that that's not the only goal. You can have an interaction that's successful and takes little time, but is unpleasant. So satisfaction is another goal."

Black believes Siri delivers that satisfaction partly through its helpful-yet- sassy tone. "It doesn't just answer questions," he says. "It has a character. It wants to name you, to know who you are. You can tell it to call you "Master" or "Darth Vader" or whatever, but it wants to call you that. It makes things a little more personal, and that's important." Faulkner also points to Siri's many handcrafted, hidden Easter eggs. For instance, if



"IT WOULDN'T BE POSSIBLE TO DO SOMETHING LIKE SIRI 10 TO 15 YEARS AGO BECAUSE YOU COULDN'T GET ENOUGH DATA TO TRAIN THE SYSTEM," SAYS ALAN W. BLACK.

More importantly, today's software agents have taken context to a level never before attempted, striving to know more about you than you know about yourself—your situation, tastes, and patterns—before running off to find exactly what it believes you want. They paradoxically expand your power by limiting its domain, collapsing infinite possibilities into a single action.

Still, Moore believes the game is far from over. "The history of this field has always been one of waves of success, followed by going into the doldrums. The joke about our success is, 'Just keep showing the same graph of escalating future returns, but don't put any dates on it.' " He has tested that theory empirically by surveying his colleagues every six years, asking them when voice recognition will hit certain milestones. But every time he resurveys them, "all the dates have moved out another six years! So the future isn't getting any closer."

"Something like Siri appears and people think, 'We've solved it!' But you can't use it in a pub or a train station," says Moore. "Then when you point out all the realities of the fantastic abilities that human beings have at holding conversations in difficult circumstances, you realize we still need to solve artificial intelligence, language, neurocomputing, and so on before we have a truly autonomous agent."

Talking To Machines

Voice recognition programs like Siri are now capable of understanding spoken commands, recognizing a conversation's context, and answering questions in a personable manner. From scheduling dinner appointments to answering trivia with wit, this book explores how digital assistants like Siri emerged at the intersection of speech recognition, AI, cloud computing, and interface design. Discover how technologies once confined to labs have transformed the way humans interact with machines—and what that means for our future.