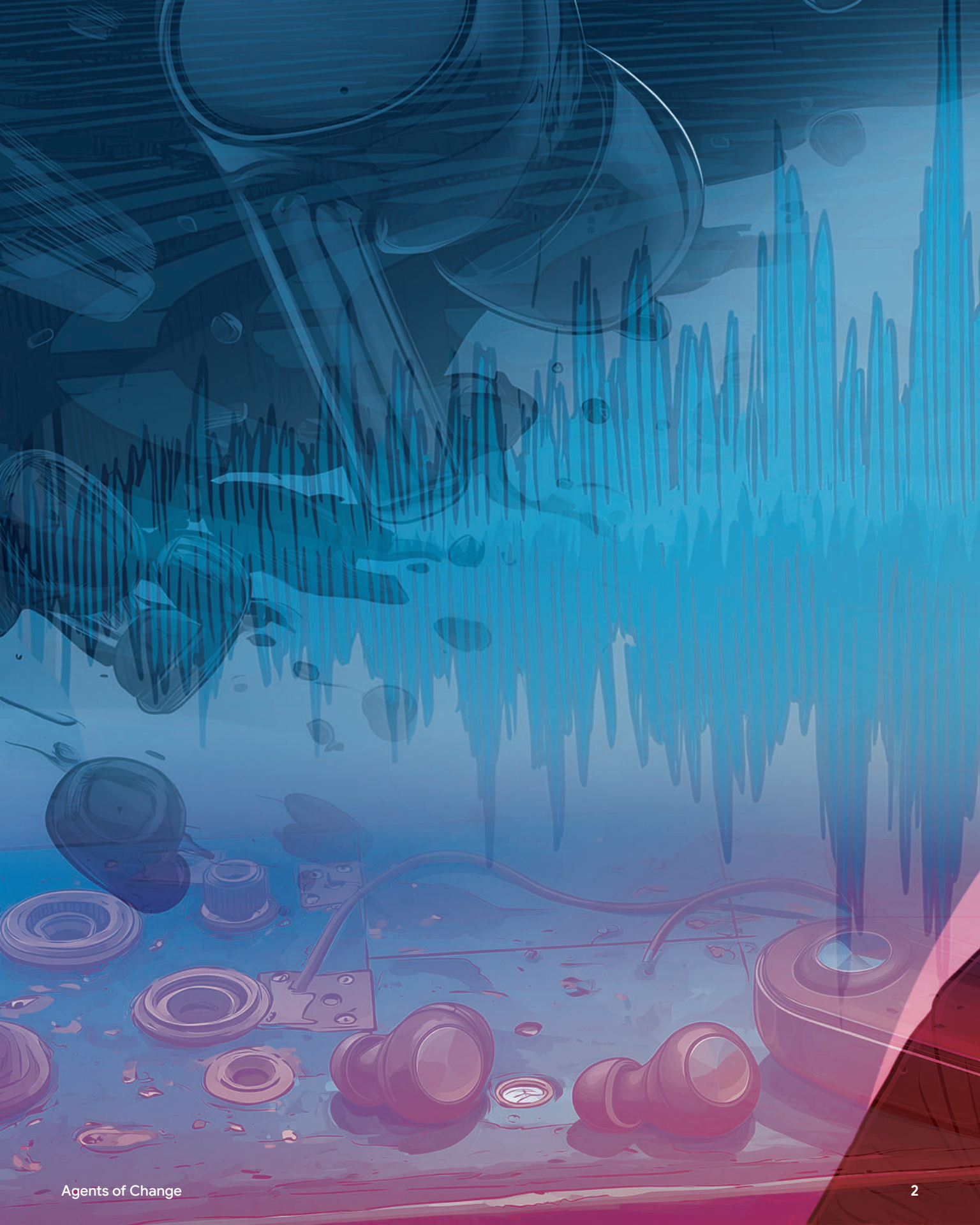




Agents of Change

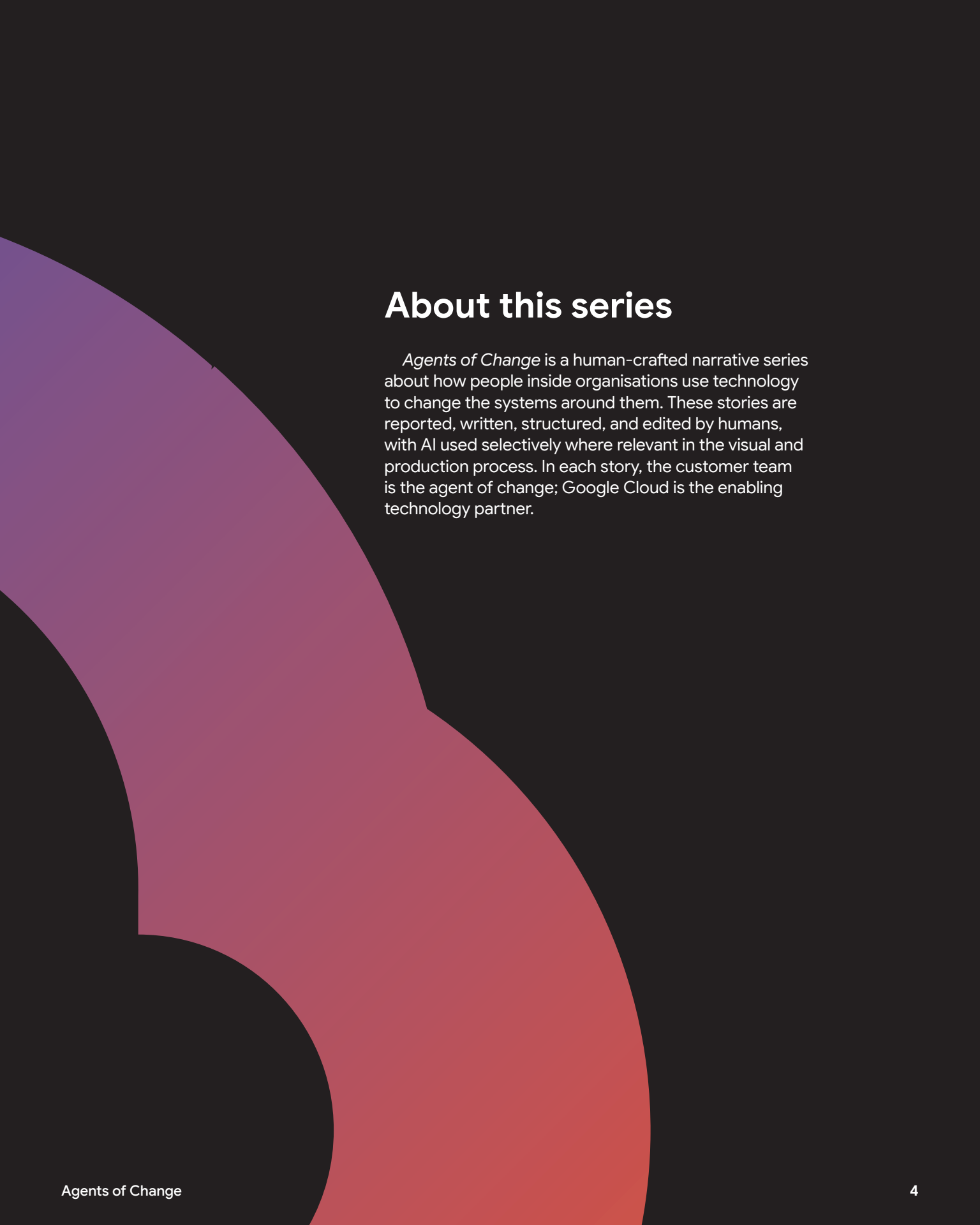


An illustration featuring a woman in the foreground on the left, looking upwards with a thoughtful expression. Behind her and to the right is a man with a slight smile, wearing a grey t-shirt. The background is a stylized cityscape with tall buildings in shades of blue and teal. Several thought bubbles are visible, some containing faint outlines of people or objects. The overall color palette transitions from cool blues at the top to warm reds and oranges at the bottom.

Speaking in your own accent: Building AI that listens to India

How Mivi used AI from Google Cloud to build multilingual conversational fluency across Indian accent variations, a capability once limited to much larger technology companies.

Chapter 2 | Agents of Change by Google Cloud



About this series

Agents of Change is a human-crafted narrative series about how people inside organisations use technology to change the systems around them. These stories are reported, written, structured, and edited by humans, with AI used selectively where relevant in the visual and production process. In each story, the customer team is the agent of change; Google Cloud is the enabling technology partner.

The Breakthrough

Mivi became an agent of change by taking on a problem that has limited voice technology in India for years: how to make conversational AI work reliably across Indian languages, accent variations, and speaking patterns at a level that feels natural to everyday users.

This was not a cosmetic challenge. In a country where people switch between languages, speak English through strong regional accents, and expect technology to understand them without forcing them into artificial speech patterns, conversational fluency is hard to build and even harder to scale.

For a consumer technology company, the gap between a system that technically works and one that people actually trust can decide whether a product is adopted or abandoned. Working with Google Cloud, Mivi set out to close that gap.

The breakthrough was not simply adding AI to an existing product, but building a more capable multilingual conversational layer, one that could handle Indian accent variation, respond more naturally, and bring a level of voice interaction previously available mainly to much larger global technology companies.

That shift mattered for two reasons. First, it changed the product experience itself: voice interaction could become more fluid, less frustrating, and more accessible to a wider base of Indian users. Second, it changed Mivi's own position in the market, allowing it to move from being a hardware and device company toward something more ambitious: a company building intelligence into the everyday interface between people and technology.

The story that follows is about what it took to get there: the technical difficulty, the pressure of building at that level, and the people who had to turn a hard engineering problem into something that could survive real-world use.

Why voice became the north star: a want is born



I THINK FASTER
THAN THIS THING
CAN KEEP UP...



IT DIDN'T BEGIN WITH A PRODUCT PLAN. IT BEGAN
WITH AN IRRITATION.



VISHWANATH TYPED QUESTIONS
THE WAY EVERYONE DID. ONE
PHRASE, ONE CORRECTION,
ONE REVISION AT A TIME.



...Mxxd. Tels.
.శావర. ఎదురు శు.
Teii.leggg.

రకే.వదిన
రామసుమిత్రే.

Tũsమ (arpgss

బిశాక్షణ !
పాఠశాల

Pkel
ekengu
bek.



THE PROBLEM
WAS THAT TYPING
SLOWED HIM DOWN.
IT MADE HIM SELF-
CONSCIOUS. IT
FORCED HIM TO
TIDY THOUGHTS
THAT WEREN'T
MEANT TO BE TIDY.

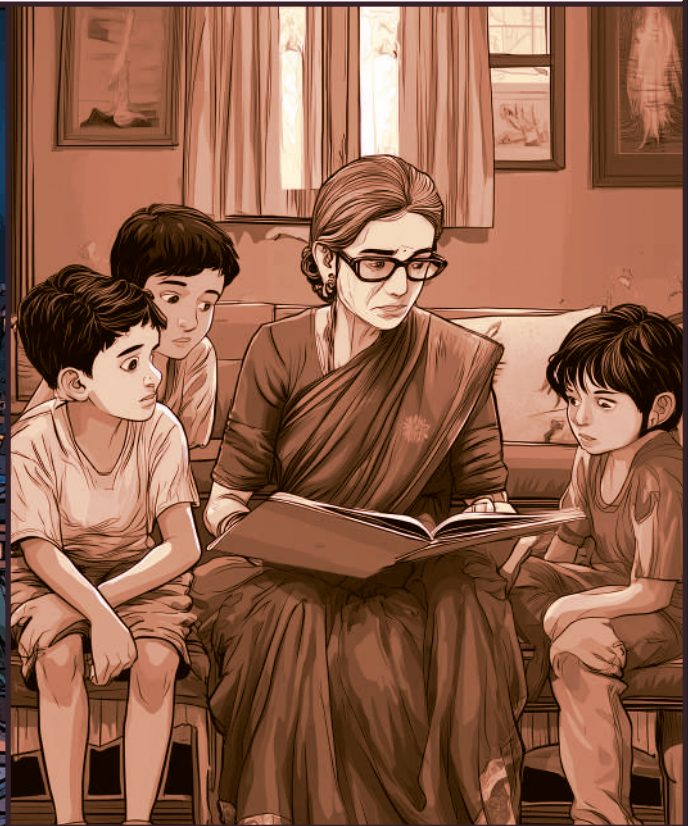
THE AI UNDERSTOOD HIM PERFECTLY WHEN HE TYPED INTO THE
COMPUTER. BUT THAT WASN'T THE PROBLEM.



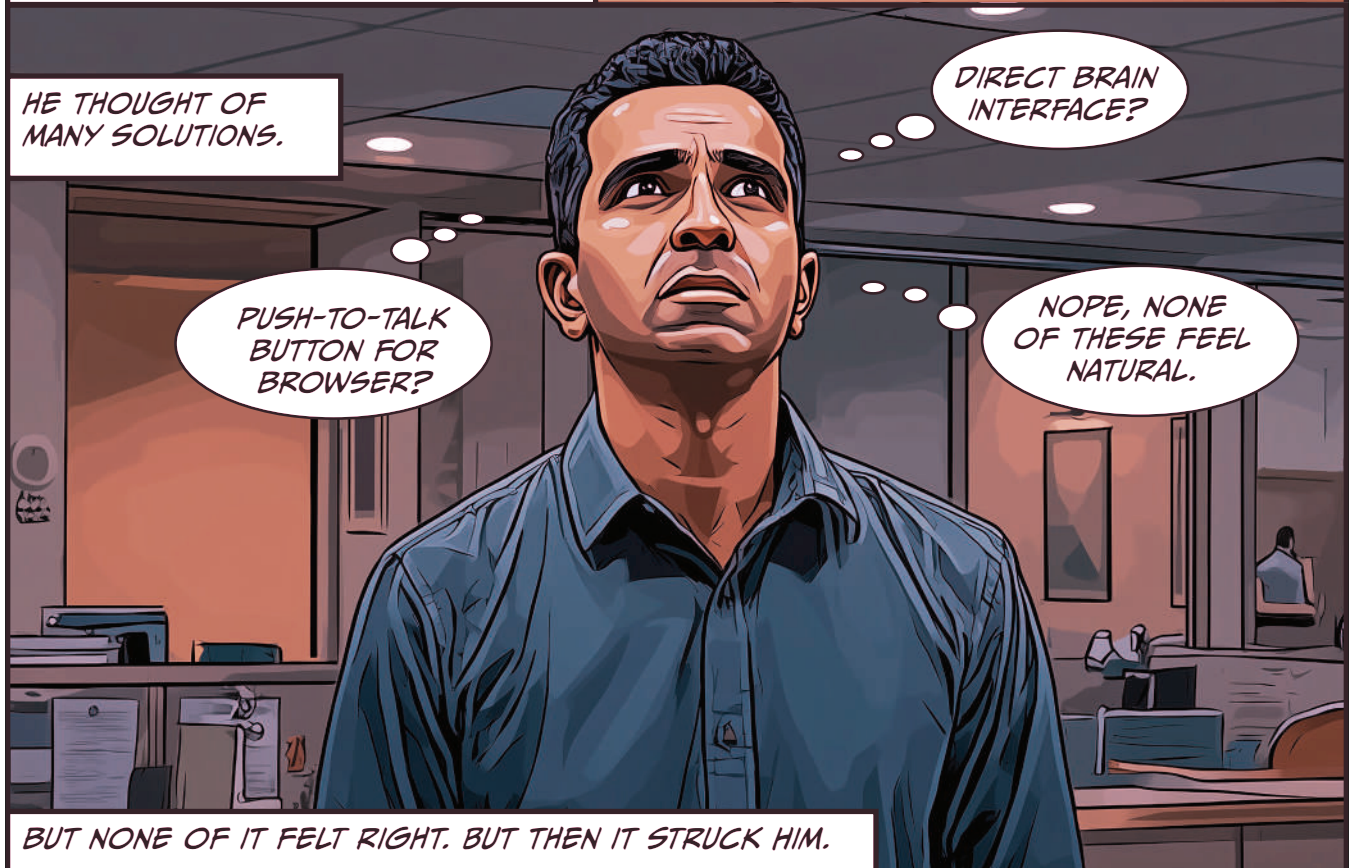
WHEN HE SPOKE, WORDS ARRIVED UNFILTERED. AT THE PACE OF HIS OWN MIND.



IF AI WAS A "KNOWLEDGE ENTITY," WHY WERE HUMANS STILL BENDING TO ITS INTERFACE?

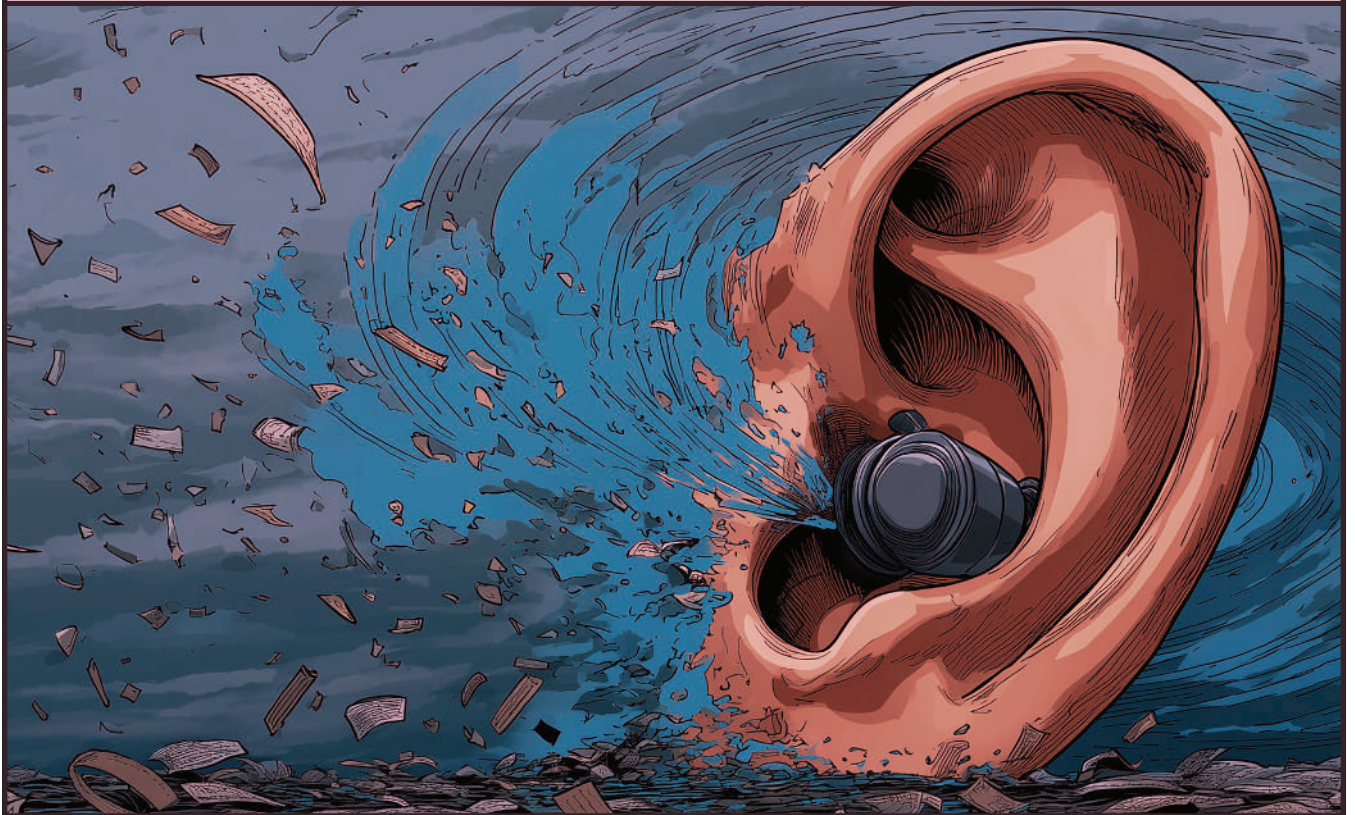


VISHWANATH RECALLED HIS CHILDHOOD. WHERE STORIES WERE OFTEN TOLD NOT TYPED. FOR MOST OF HUMAN HISTORY, KNOWLEDGE MOVED BY VOICE.



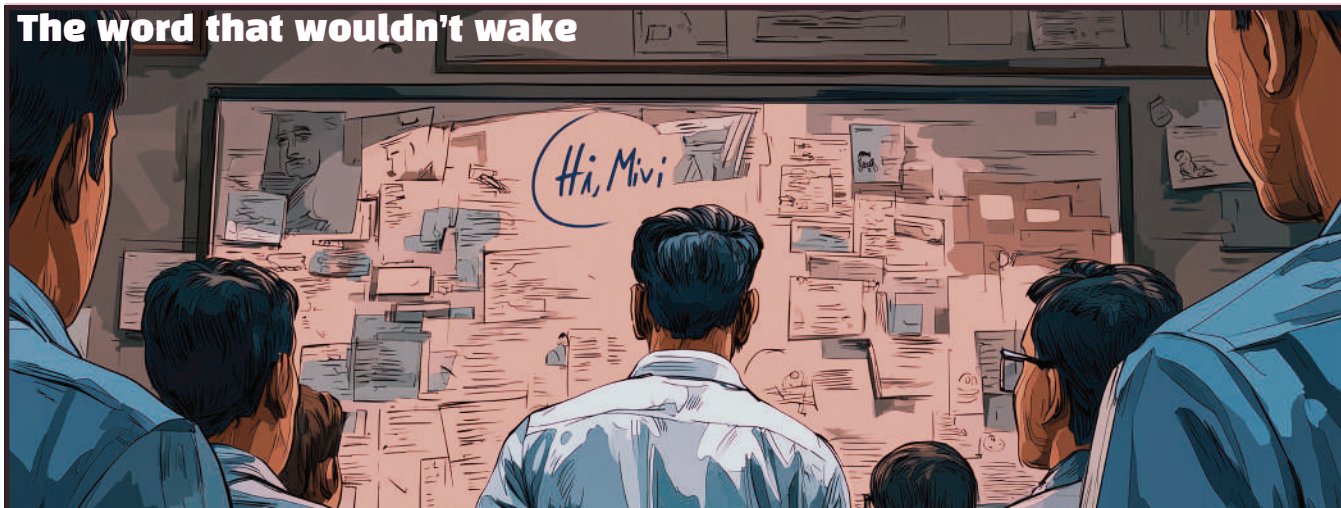


NOT 'LET'S ADD VOICE TO EARBUDS,' BUT 'LET'S BUILD THE HUMAN-AI INTERFACE AROUND VOICE.'

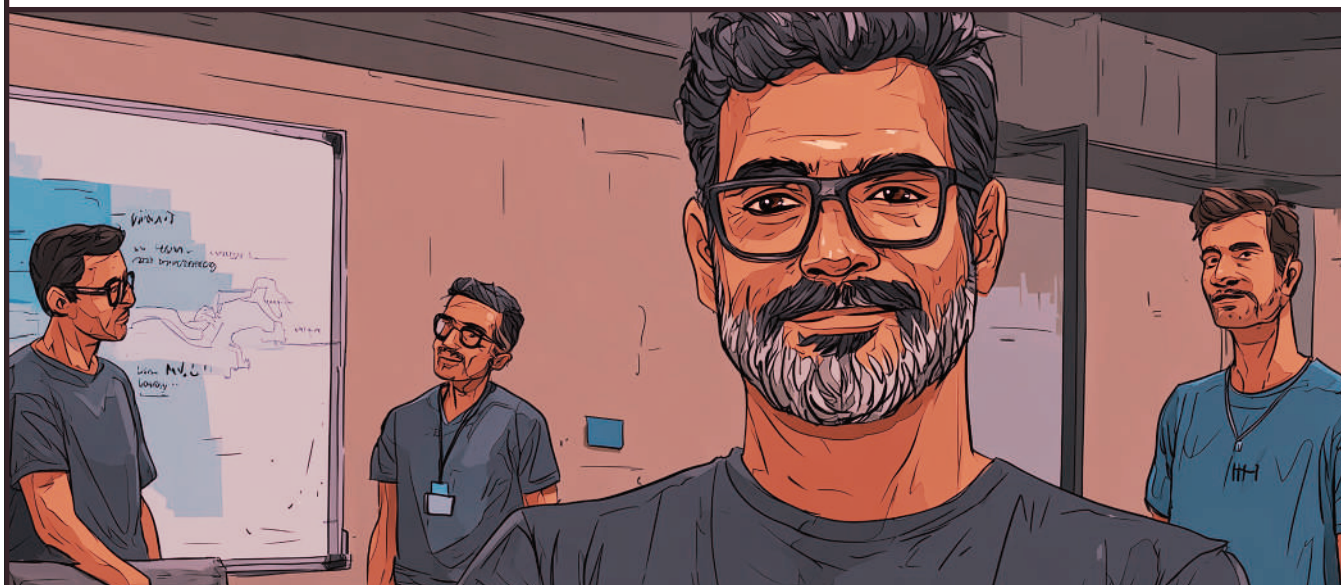


IF AI WAS GOING TO BECOME A PRESENCE IN DAILY LIFE, IT HAD TO ENTER THROUGH THE EAR BEFORE IT ENTERED THROUGH THE EYE.

The word that wouldn't wake



THE WHITEBOARD CARRIED BIGGER AMBITIONS. BUT TWO SYLLABLES ALWAYS CAME BACK DARKER: HI MIVI, THE WAKE WORD.



THEY THOUGHT IT WOULD BE THE EASY PART. THE PRELUDE BEFORE THE REAL WORK BEGAN.

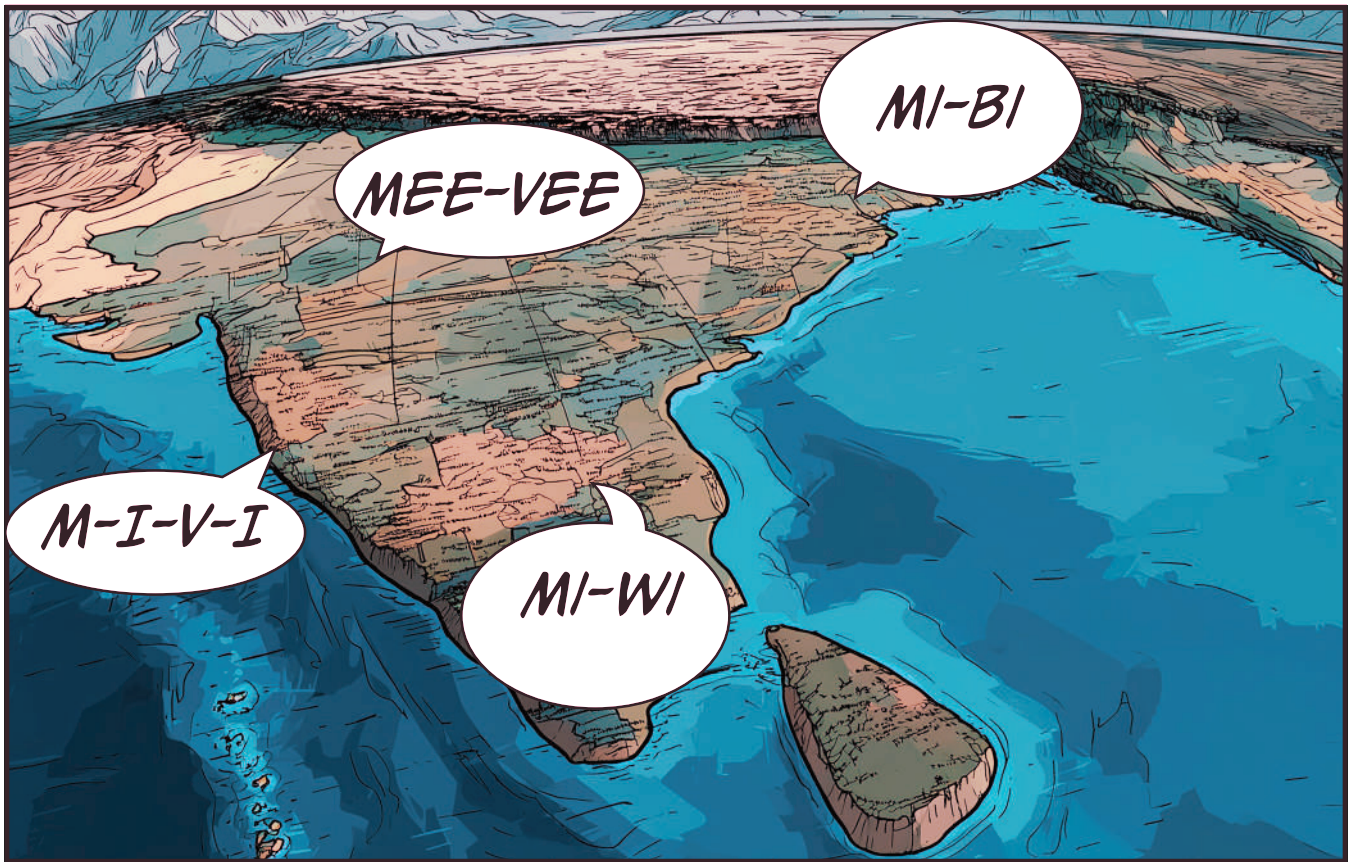


HI
MIVI.

NOTHING
HAPPENS.

THIS TIME IT
WAKES INSTANTLY.

INSTEAD, IT BECAME THE TEST THAT REFUSED TO END



AND INDIA POSED UNIQUE CHALLENGES. INDIA DOES NOT SAY "MIVI" ONE WAY.



"HI MIVI WAS
OUR HARDEST
LESSON." —
MIDHULA



THEY BEGAN TO SOLVE THE PROBLEM AT THEIR OWN FACTORY FLOOR, A MICROCOSM OF INDIA'S DIVERSITY.

EVERYBODY RECORDED. EVERYBODY SPOKE IN THEIR OWN VOICE. AND IT ALMOST WORKED.



FOR A MOMENT IT LOOKED SOLVED. INTERNAL TESTS PASSED. THE DEVICE FELT ALERT.

BUT WHEN THE FIELD UNITS WENT OUT, THE MISSES BEGAN.



IT WASN'T ONE ACCENT. IT WAS A SCATTER OF MISSES WITH NO CLEAR EXPLANATION.



SO THEY WIDENED THE CIRCLE. DISTRIBUTORS, PARTNERS, STORE STAFF, ANYONE WITHIN REACH WERE ASKED TO PARTICIPATE.



THE RECORDED VOICE DATA ADDED UP. UNEVEN. NOISY. BUT IT HELPED TRAIN THE AI MODELS.

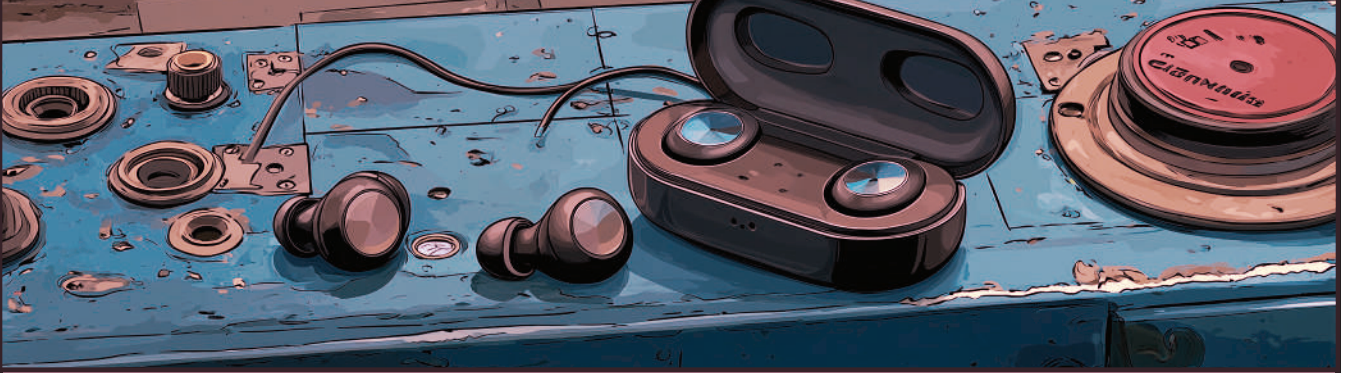


EVEN THAT WASN'T ENOUGH. THERE WERE VOICES THE SYSTEM SIMPLY WOULD NOT HEAR. TRAINING ALONE COULD NOT UNTIE THE KNOT.



THEY KEPT THE TWO WORDS ON THE BOARD. NOT FOR LACK OF VISION, BUT BECAUSE NOTHING ELSE MATTERED IF THE BUD WOULD NOT WAKE WHEN AN INDIAN VOICE CALLED IT BY NAME.

The chip that said no



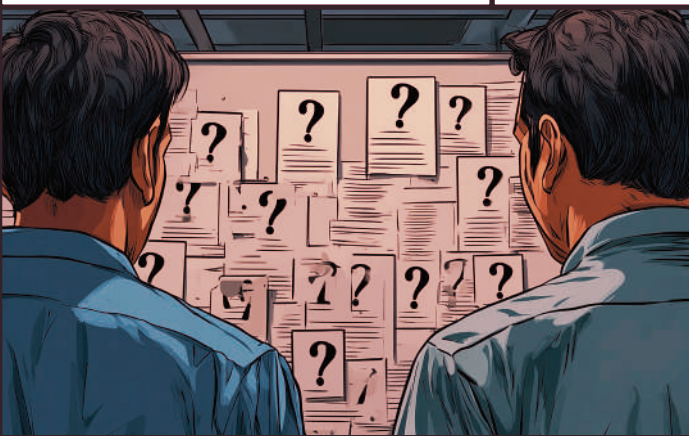
EVEN WHEN THE EARS WERE READY, PHYSICS HAD ITS SAY.



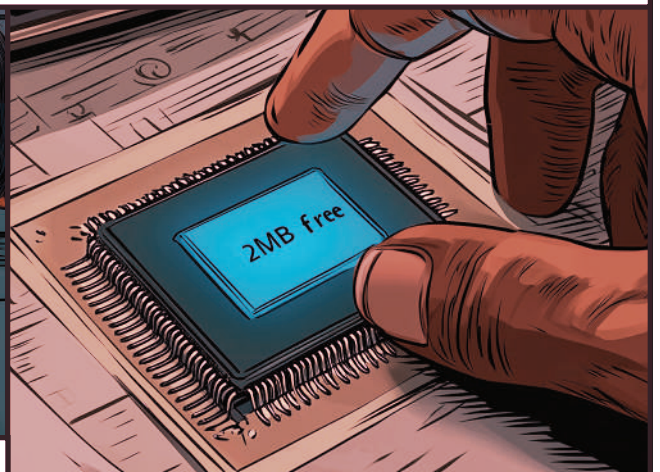
*MINI'S BUDS HAD TO WORK
WITH WHATEVER PHONE INDIA
CARRIED
EACH BRAND MANAGES
BACKGROUND TASKS IN ITS
OWN PRIVATE WAY.*



*A TRIGGER THAT WOKE ON A PIXEL PHONE COULD DIE
QUIETLY ON ANOTHER PHONE.*



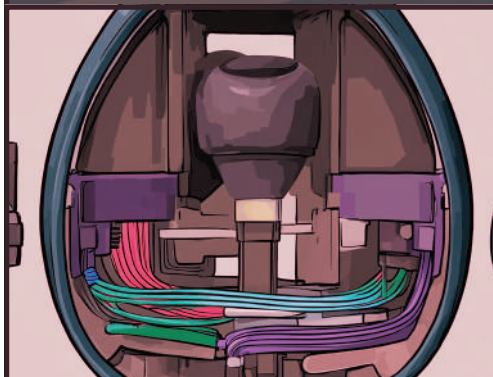
*SOME RULES WERE WHITELISTED. OTHERS
WEREN'T WRITTEN ANYWHERE.*



AND THEN THERE WAS THE CHIP.



THE IC MEMORY IS ONLY 2 MB... OUR FIRST MODEL WAS 10 MB.



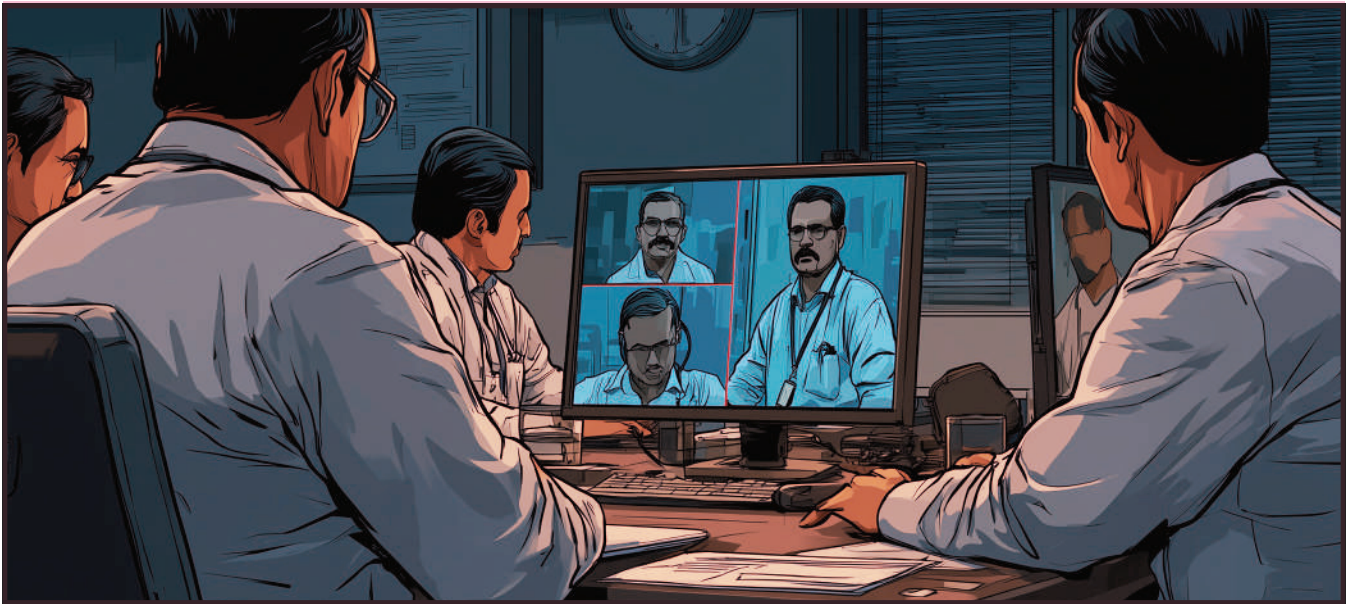
YOU CAN ADD EXTERNAL MEMORY. BUT ONLY IF THERE'S SPACE. THERE WASN'T.



SO THEY SHAVED. AND SHAVED. UNTIL THE MODEL FIT LIKE A JACKET PULLED TIGHT AT THE SEAMS.



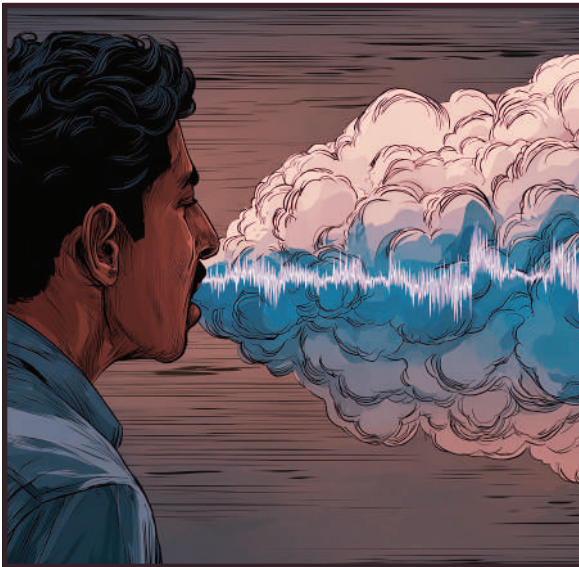
EARLY EXPERIMENTS COLLECTED TEN COMMANDS. ACCENTS MANGLED HALF OF THEM. THEY MADE A RUTHLESS CHOICE: SHIP WITH ONE RELIABLE WAKE PHRASE. ONLY THAT.



THEY TUNED MEMORY, COMPUTE, AND HOOKS WITH THE SAME VENDORS USED BY PIXEL AND GALAXY. BUT THAT WASN'T ENOUGH.



EVEN WHEN THE MATH FIT, INDIA'S SOUNDSCAPE WAS EXTREMELY COMPLEX.



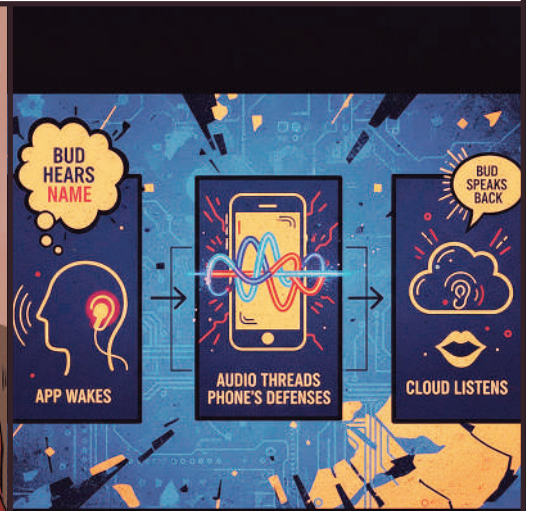
VOICES HUMANS UNDERSTOOD EFFORTLESSLY MADE THE CLOUD STUMBLE.



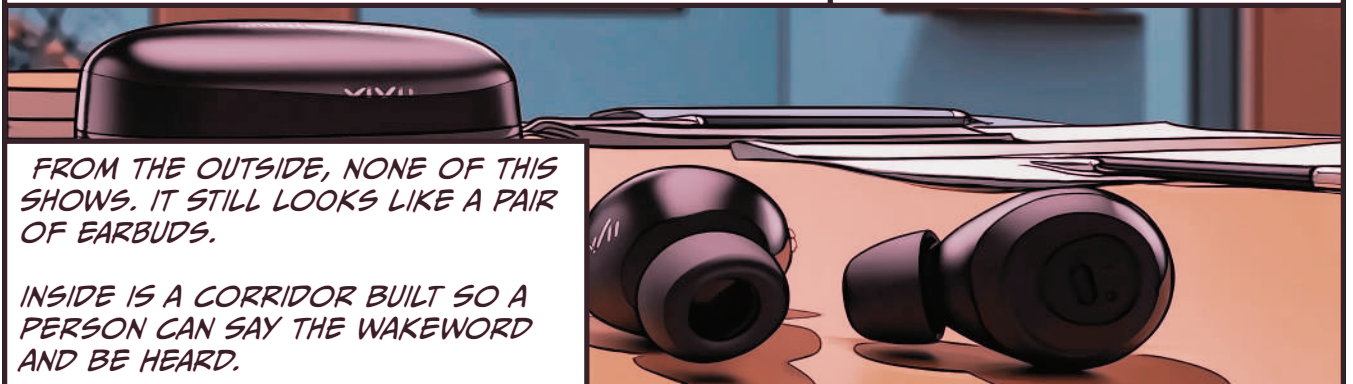
A CUSTOM NOISE-SUPPRESSION MODEL CLEANED AUDIO TWICE. ONCE IN THE BUD, ONCE IN THE APP.



SPEAKERS WHO HAD BEEN INVISIBLE TO THE SYSTEM FINALLY BEGAN TO LAND.



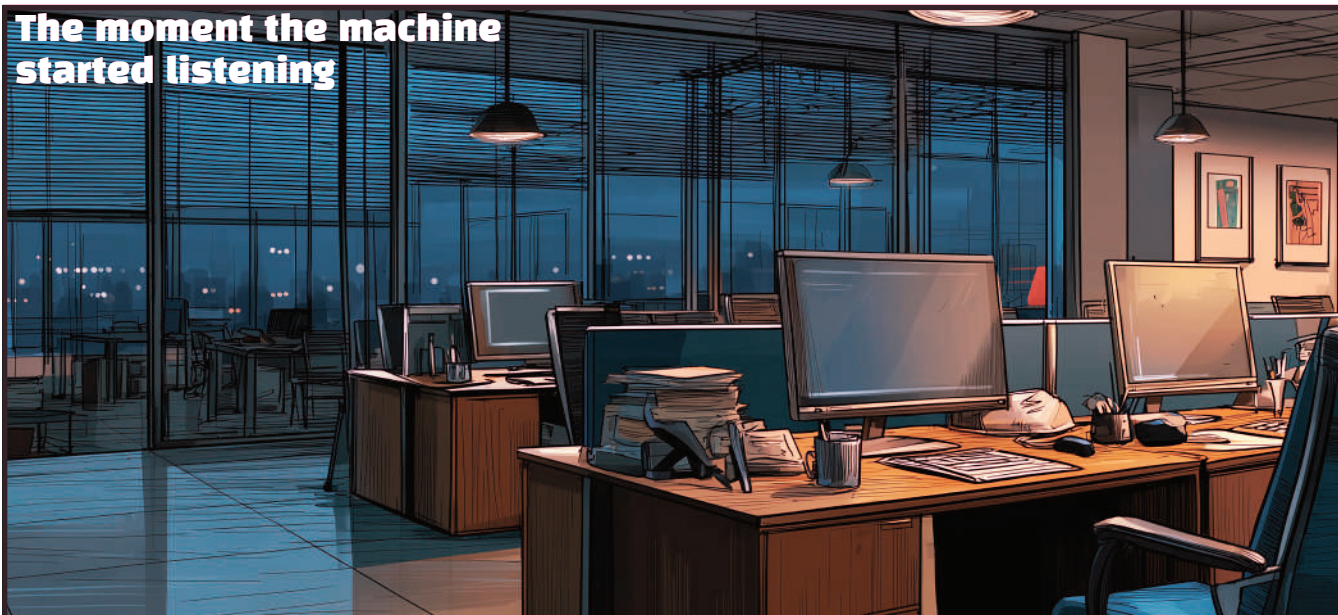
END TO END, THE PATH BECAME A RELAY BUILT METER BY METER.



FROM THE OUTSIDE, NONE OF THIS SHOWS. IT STILL LOOKS LIKE A PAIR OF EARBUDS.

INSIDE IS A CORRIDOR BUILT SO A PERSON CAN SAY THE WAKEWORD AND BE HEARD.

**The moment the machine
started listening**



IT DIDN'T ARRIVE WITH FANFARE. NO ALL-HANDS. NO CHAMPAGNE.



*MAYBE TRY— NO
WAIT...*

*OKAY BUT...
WHAT IF...*

*SUPPOSE IT
HAD TO...*

*JUST A LATE HOUR AND A HABIT
OF PACING.*

*HE SPOKE THE WAY PEOPLE DO WHEN THEY'RE
ALONE. SENTENCES THAT WANDERED, THOUGHTS
THAT CHANGED LANES MID-SENTENCE.*

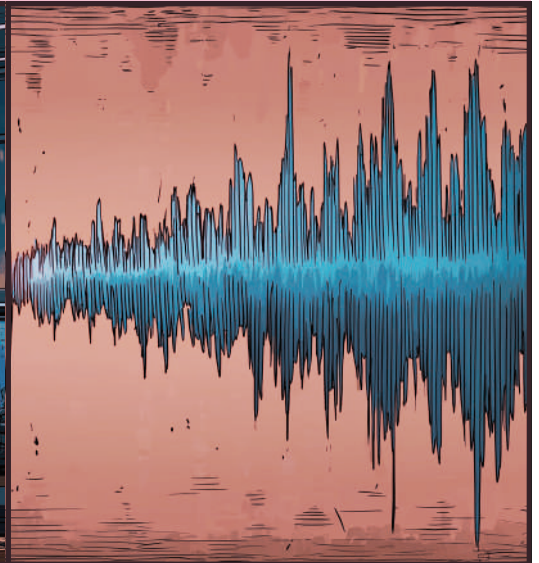


*HERE'S WHAT I
THINK...*

THE REPLY CAME BACK AT ONCE. CLEAR, LEVEL, UNHURRIED.



HE TRIED AGAIN. NO SLOWING DOWN. NO CHOOSING SIMPLER WORDS FOR THE MACHINE'S SAKE.

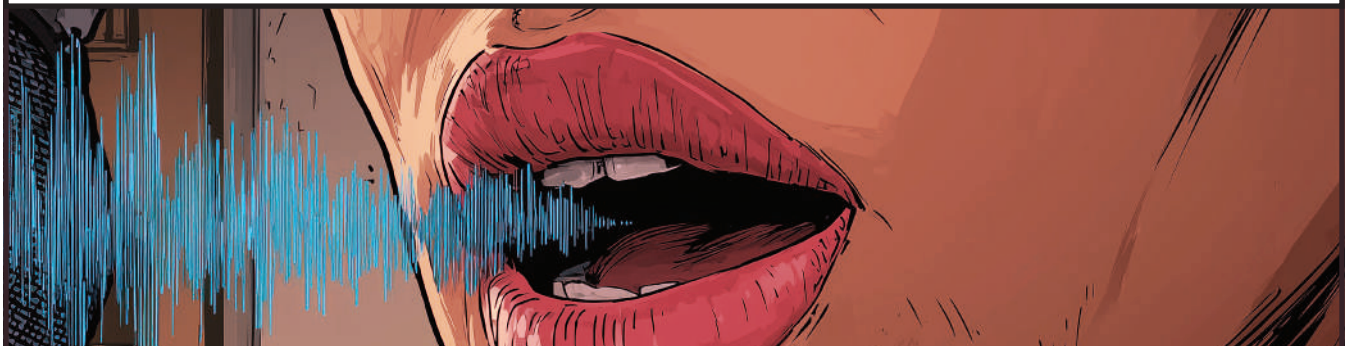


THE EARBUD ANSWERED IN THE SAME EVEN TONE.

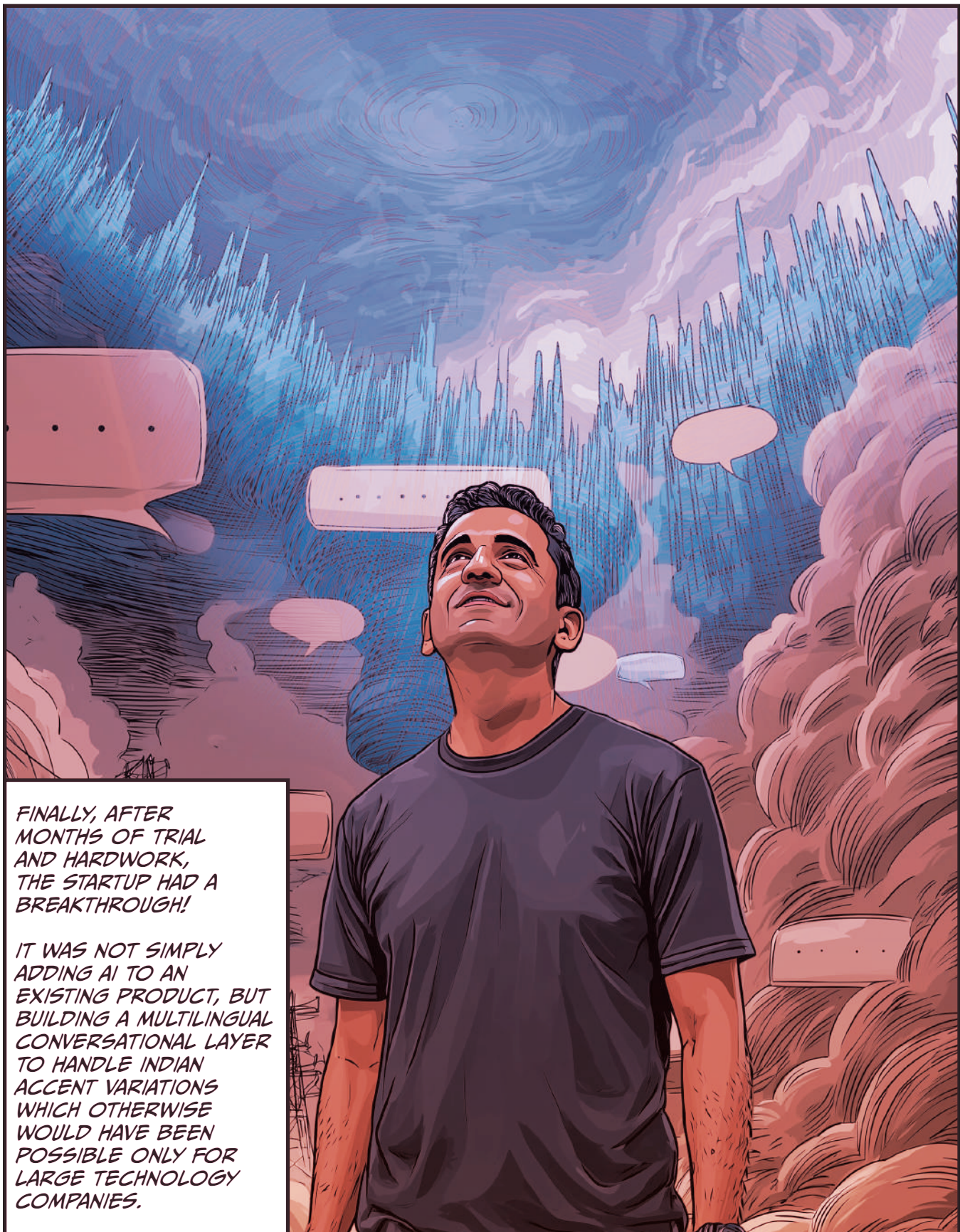


YOU
FORGET
YOU'RE
TALKING TO
A MACHINE.

WITHIN MINUTES, HE STOPPED NOTICING HE WAS TALKING TO A DEVICE.



PART OF THE EASE CAME FROM THE VOICE. NO LONGER SOFTWARE, BUT A PRESENCE TRYING NOT TO INTRUDE.

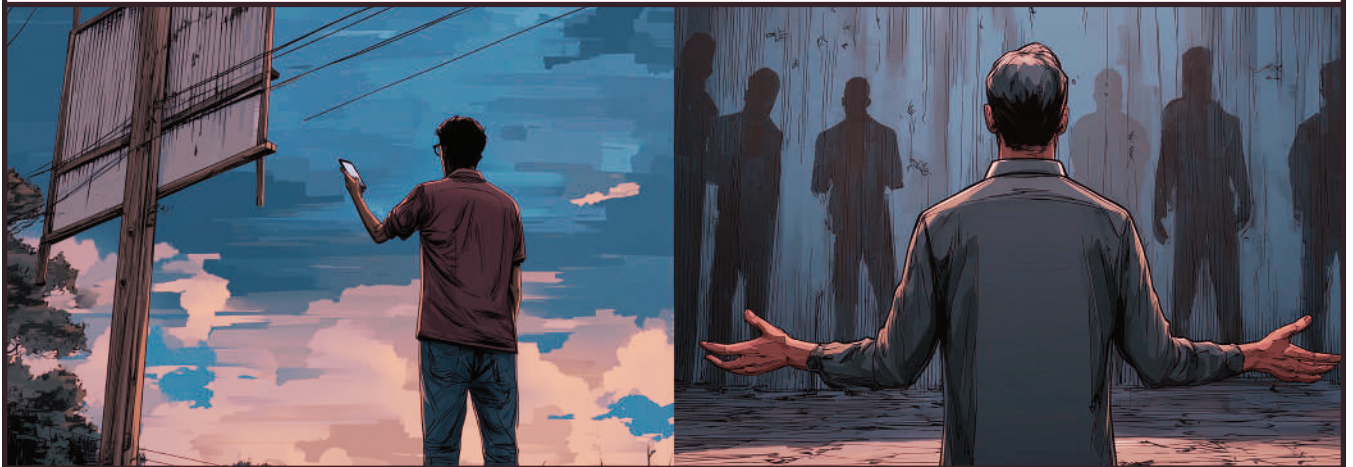


FINALLY, AFTER MONTHS OF TRIAL AND HARDWORK, THE STARTUP HAD A BREAKTHROUGH!

IT WAS NOT SIMPLY ADDING AI TO AN EXISTING PRODUCT, BUT BUILDING A MULTILINGUAL CONVERSATIONAL LAYER TO HANDLE INDIAN ACCENT VARIATIONS WHICH OTHERWISE WOULD HAVE BEEN POSSIBLE ONLY FOR LARGE TECHNOLOGY COMPANIES.



WHEN THE DISTANCE BETWEEN THOUGHT AND RESPONSE SHRANK, FRICTION DISAPPEARED.



IT MARKED A BOUNDARY: ON ONE SIDE, THE USER IS "USING" A TOOL; ON THE OTHER, THEY ARE SIMPLY TALKING.

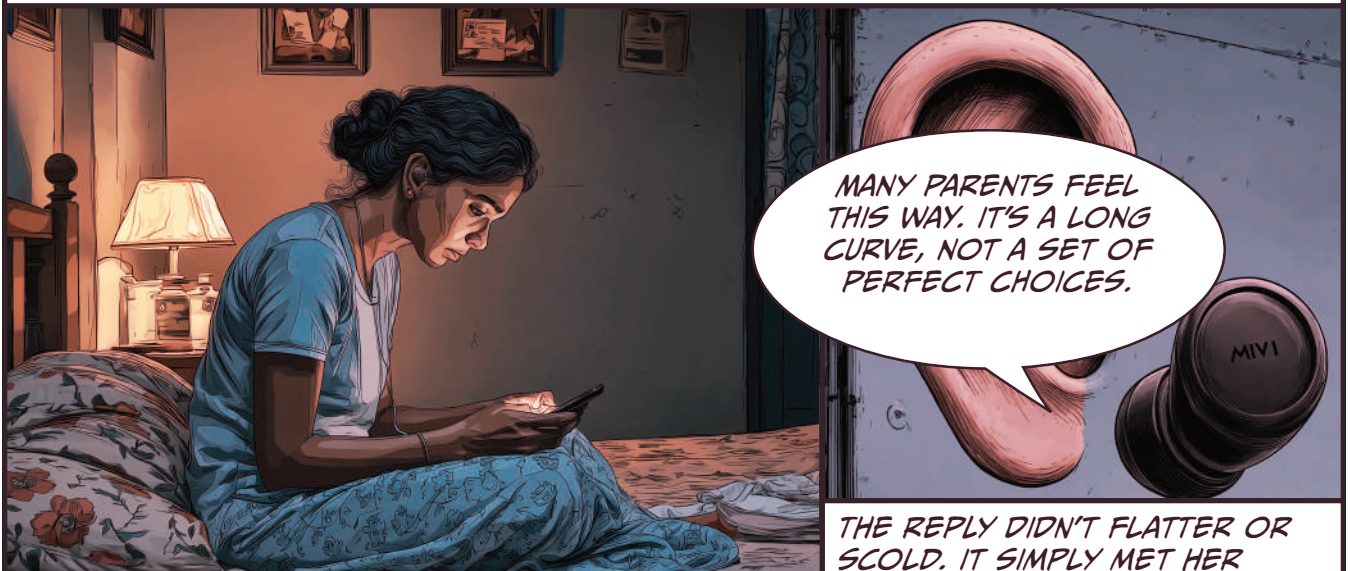


THE "TOOL" HAD BECOME SUBTLE ENOUGH TO DISAPPEAR.

The first listeners



WHEN IT FINALLY STARTED WORKING, THE FIRST STORIES THAT MATTERED WEREN'T ABOUT KNOWLEDGE. THEY WERE ABOUT BEING HEARD.



MANY PARENTS FEEL THIS WAY. IT'S A LONG CURVE, NOT A SET OF PERFECT CHOICES.

A MOTHER SPOKE ABOUT RAISING A CHILD AND THE FEAR SHE WASN'T GETTING ANY OF IT RIGHT.

THE REPLY DIDN'T FLATTER OR SCOLD. IT SIMPLY MET HER WITHOUT JUDGMENT. SHE FELT LESS ALONE.



ORIGAMI, MAYBE TRY IT BEFORE THE TRIP?

ANOTHER TESTER ASKED FOR A TRIP PLAN AND GOT EXACTLY WHAT SHE EXPECTED.

WHAT SURPRISED HER WASN'T THE IDEA, BUT THAT THE MACHINE REMEMBERED.



I BROKE THE
DIET TODAY.

YOU LIKE HEAT.
MAKE IT TWO
TABLESPOONS
OF CHILI.

I'LL WAIT.

LET'S PICK
IT UP FROM
TOMORROW.
HERE'S WHERE
WE LEFT OFF.

THAT LINE ISN'T
OURS. IT BELONGS
TO THE USER.

HOW FAR SHOULD
IT GO INTO THE
PERSONAL?

WILL YOU BE
OKAY STEPPING
OUT AT THREE
DEGREES?

WHAT'S
MOSCOW LIKE IN
OCTOBER?

THE MEMORIES WEREN'T THEATRICAL BUT DOMESTIC—SMALL MOMENTS IN KITCHENS, COLD OCTOBERS REMEMBERED, QUIET CONFESSIONS WHISPERED WITHOUT JUDGMENT. SOME PEOPLE WANTED THE MACHINE IN EVERYTHING, OTHERS AVOIDED IT ENTIRELY, AND YET IT LEARNED TO SERVE BOTH. IN THE END, THE POINT WAS NEVER SPECTACLE, BUT THE ORDINARY, EVERYDAY TEXTURE OF EARLY USE.



FINALLY, THE TEAM HAD IT! AN AI THAT KNOWS YOU, CARES FOR YOU, REMEMBERS YOU, AND IS THERE FOR YOU.

The final showdown: The skeptic in the room

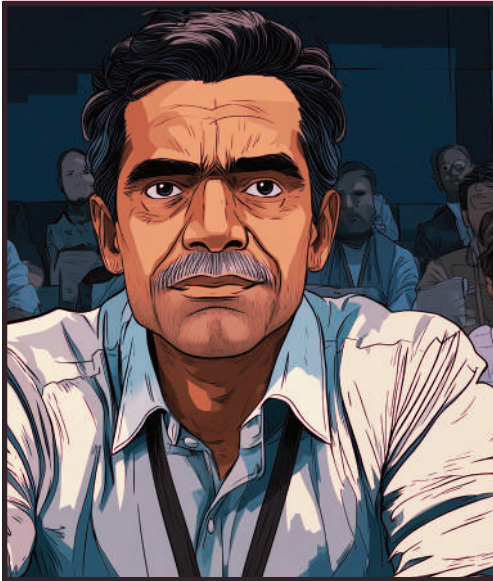


THE PITCH FROM MIVI HAD LANDED IN THE JOURNALISTS' INBOX LIKE DOZENS OF OTHER PITCHES. IT SOUNDED LIKE A STRETCH. MAYBE EVEN A GIMMICK.

THE JOURNALIST DECIDED TO GO ANYWAY. MORE OUT OF OBLIGATION THAN INTEREST.



MIDHULA AND VISHWANATH, THE TWO FOUNDERS OF MIVI WERE HERE TO DO THE UNTHINKABLE. PROVE THE PRODUCT ACTUALLY WORKED.



MIDHULA HAD LEARNED TO READ THAT LOOK. POLITENESS MASKING DISBELIEF.

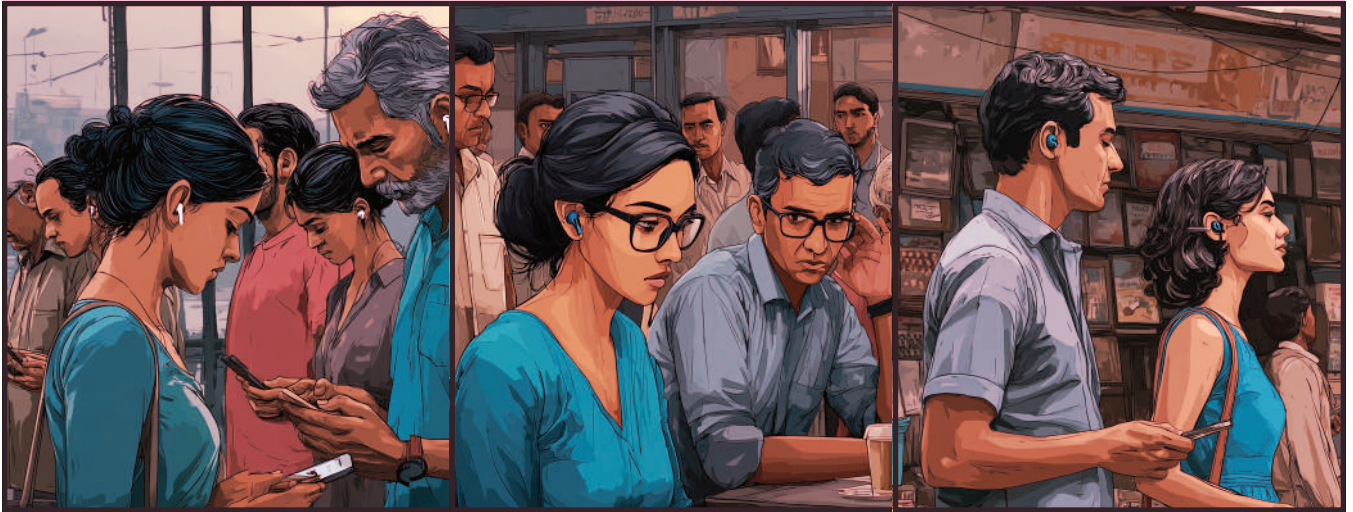


AI INSIDE THE
EARBUDS... EIGHT
LANGUAGES... NO
NEED TO REACH FOR
THE PHONE...

NOTHING IN THE ROOM HAD CHANGED YET.
THE JOURNALIST GLANCES AT HIS WATCH.



WHY DON'T
YOU JUST
TRY IT?



ACROSS INDIA, THE PATTERN HAD BEEN THE SAME: SKEPTICISM FIRST, BELIEF ONLY AFTER THE DEMO.



IN DELHI, THE JOURNALIST SLIPPED THEM IN AS IF TO HUMOR THE FOUNDERS.

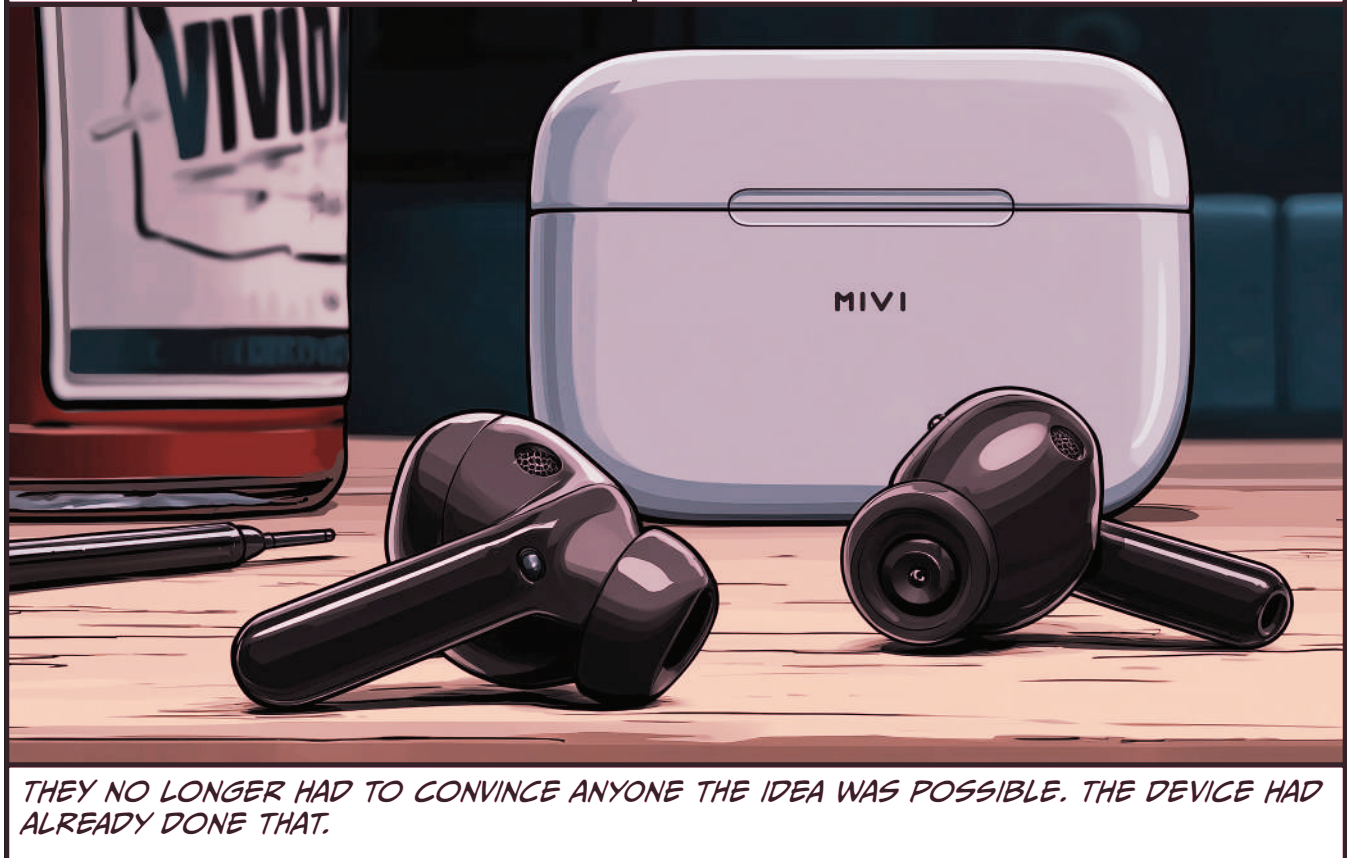
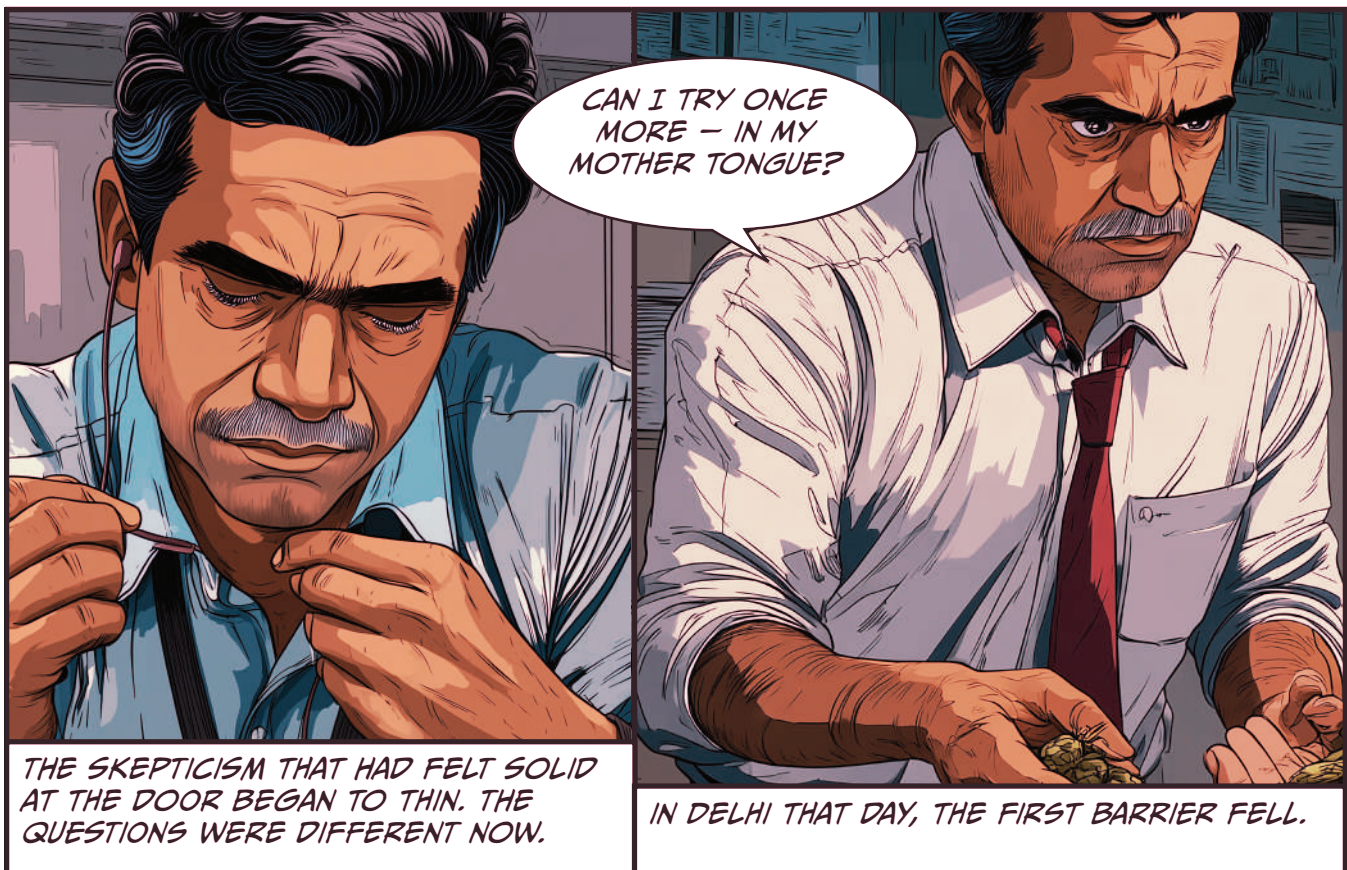
STANDARD QUESTIONS. STANDARD ANSWERS. NO ROBOTIC EDGE.



THEN HE SWITCHED TO TAMIL. FAST, NATURAL, UNWARNED.

THE REPLY LANDED CLEAN. FLUENT. UNHESITATING.

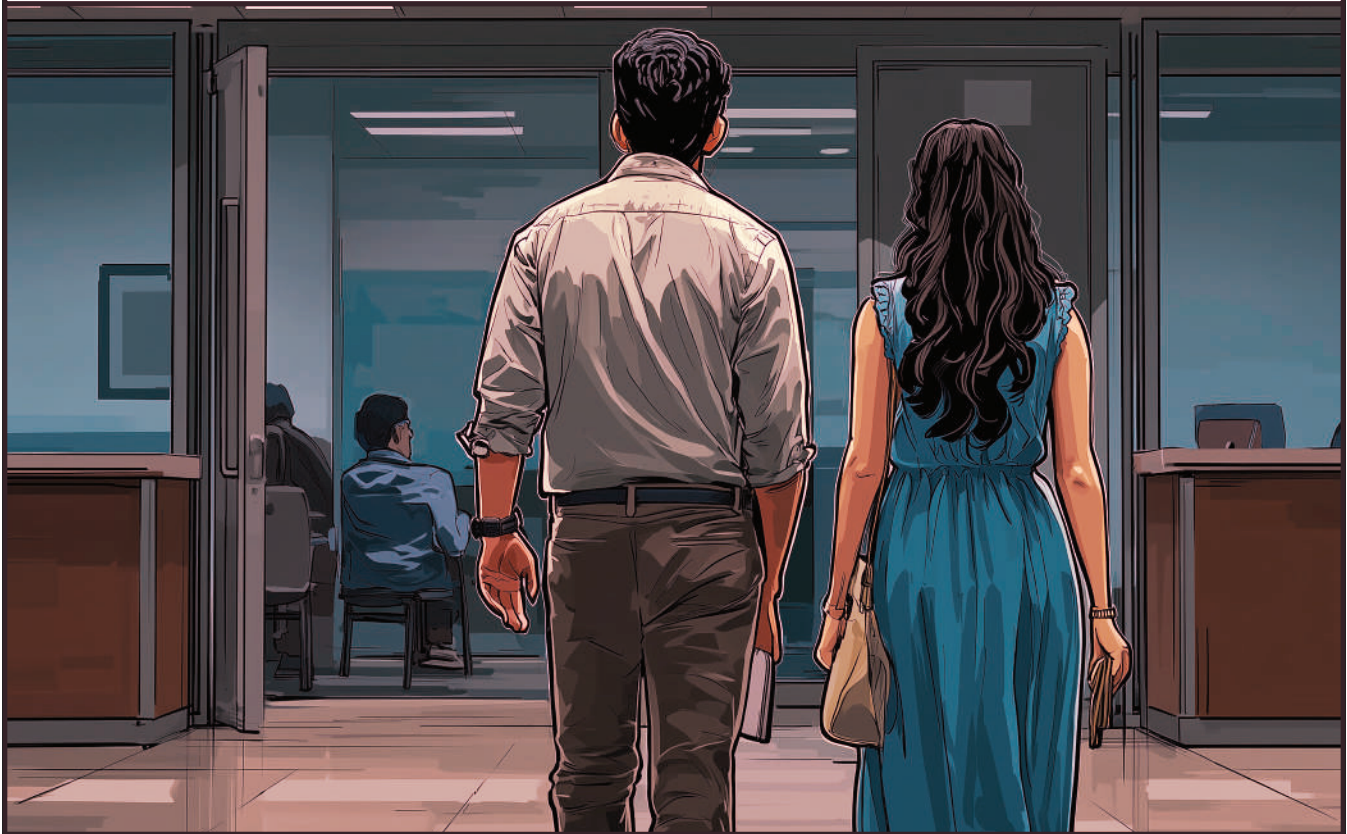
THEN MARATHI. ANOTHER SEAMLESS ANSWER.



**The big vision:
What they are really building**



ASK THEM WHAT THEY'RE MAKING AND THEY WILL NAME A PRODUCT.



THEY ARE BUILDING FOR THE SMALL, MISUSED STRETCHES BETWEEN THINGS.



THEY ARE BUILDING A VOICE
ENABLED COMPANION.



ORIGAMI?

WHICH NOT ONLY
REMEMBERS, BUT IS NOT
JUDGEMENTAL.



THEY ARE TURNING
CONSUMER ELECTRONICS
INTO CARRIERS OF
INTELLIGENCE.



IT WASN'T JUST AN API SUBSCRIPTION TO AI FROM GOOGLE CLOUD. IT WAS A TRUE
COLLABORATION.



THEY ARE BUILDING A WAY FOR A PERSON IN INDIA TO BE HEARD BY A MACHINE WITHOUT CHANGING
THEMSELVES TO FIT IT. THEY ARE BUILDING THE FUTURE OF AMBIENT COMPUTING FOR INDIA.

English science fiction writer Arthur C Clarke wrote: Any sufficiently advanced technology is indistinguishable from magic. Where AI is today would have been magic to the average person from a few years ago.

The shift from typing to talking looks simple: a pair of earbuds, a natural question, a thought delivered unfiltered. But that simplicity only holds because of what came before it: the “slow work under the hood,” the months of Mivi and Google engineers poring over spectrograms, and the choice to build for the small, misused stretches of time between things.

It wasn’t just an API subscription to AI from Google Cloud; it was a true collaboration to turn everyday Indian appliances into carriers of intelligence. Because when technology finally aligns with the pace of the human mind, it stops being a machine we have to master and becomes a companion that listens without judgment. It becomes magic.

Coming up Next

The orchestrators: Breaking the enterprise paradox

