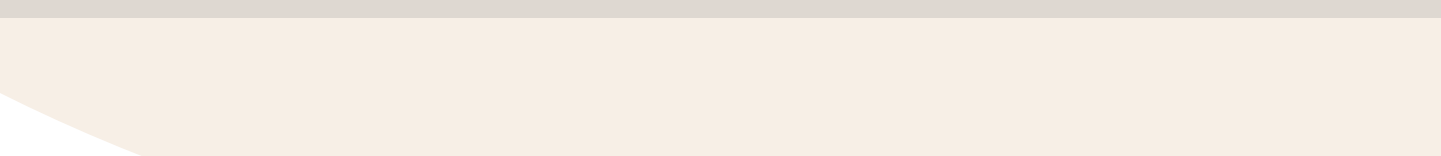


MARKET · DATA · TRUST

What is Conversational AI?

How it works, examples, and more

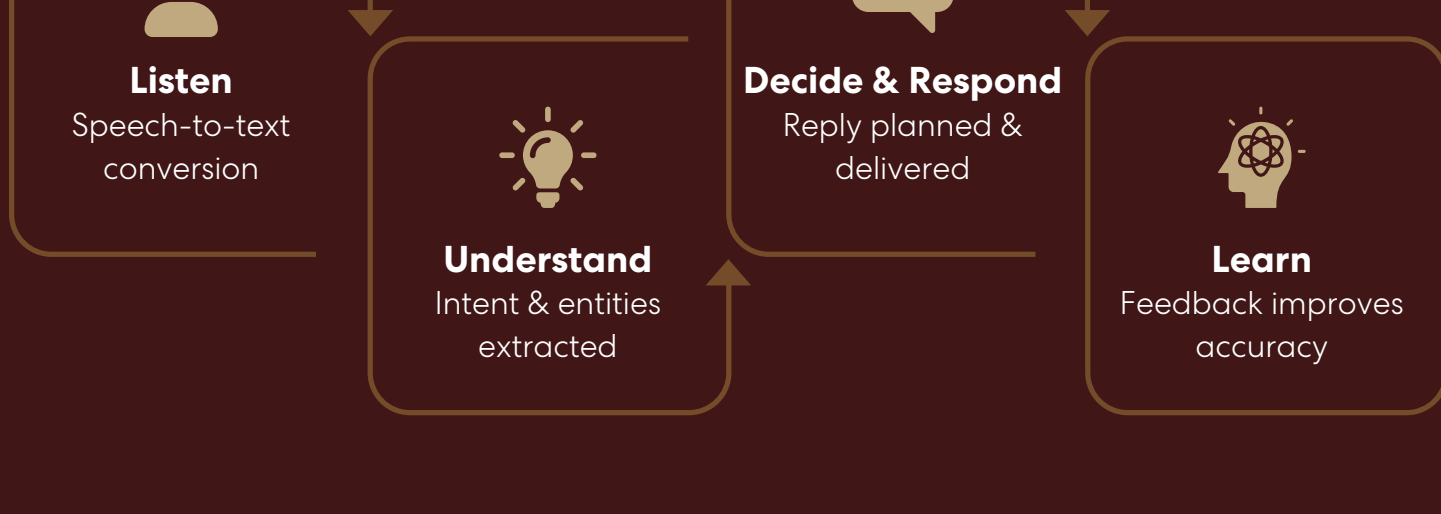
How the technology works, what it takes to collect & label the data behind it, and how to keep that data trustworthy and compliant, from raw speech to a production-ready model.



WHAT IS CONVERSATIONAL AI?

Machines that Understand & Respond Naturally

Chatbots, virtual agents, and voice assistants that combine speech recognition, language understanding, and speech generation to hold a natural, back-&-forth conversation & get smarter each time they're used.

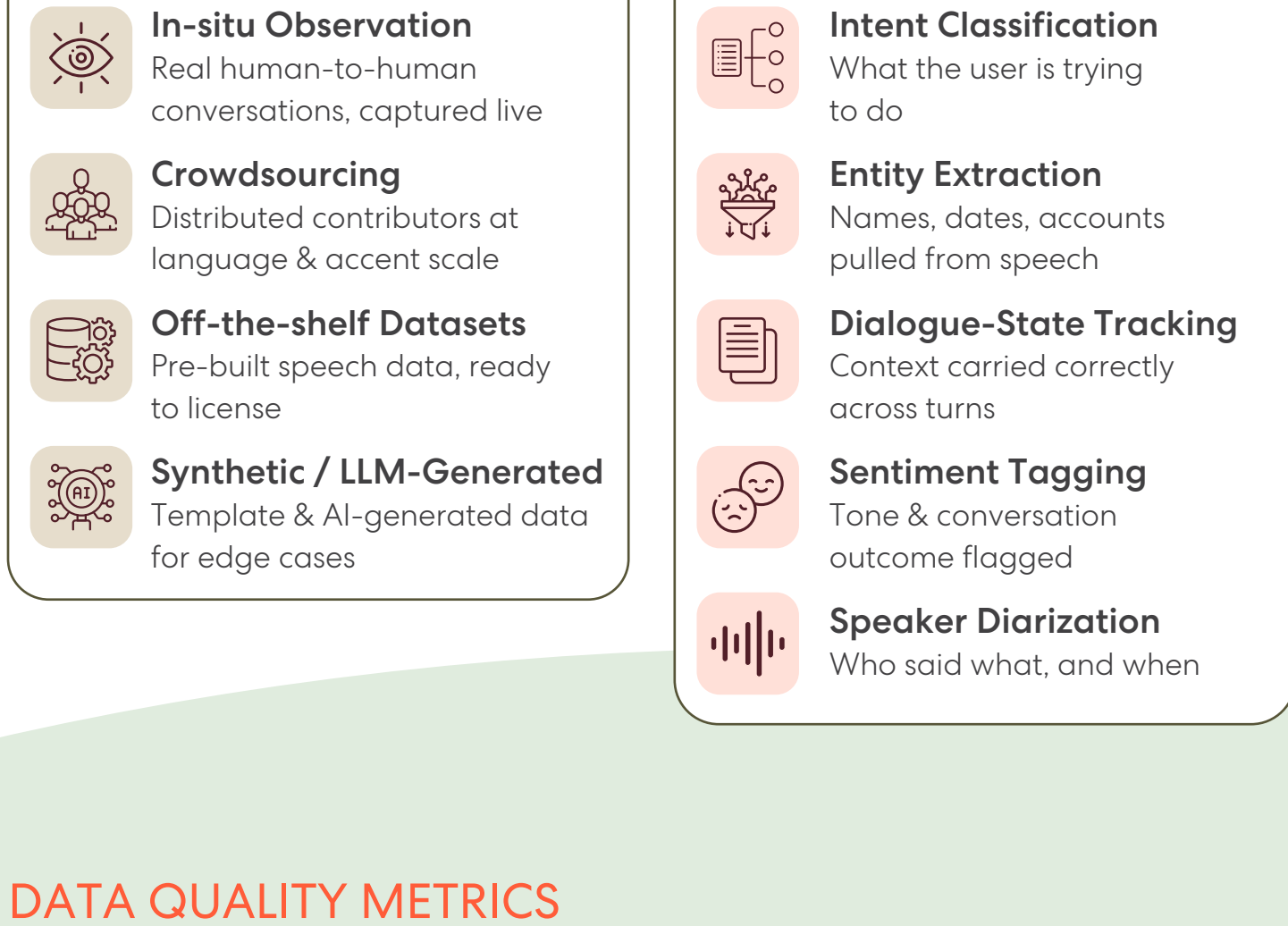


- | | |
|--|--|
| Voice Assistants
Alexa, Google Home & smart speakers | Support Chatbots
FAQ & customer service automation |
| Banking Bots
Balance checks, fraud alerts, KYC | Healthcare Assistants
Scheduling, triage & patient support |
| HR Bots
Onboarding, leave requests, policy Q&A | IoT Devices
Wake-word & far-field voice control |

DATA COLLECTION & ANNOTATION

How Raw Speech Becomes Training Data

Getting from a recording to a model-ready dataset takes two passes: choosing how the speech is gathered, then labeling it so a model can learn from it.



DATA QUALITY METRICS

What "Good Data" is Measured By

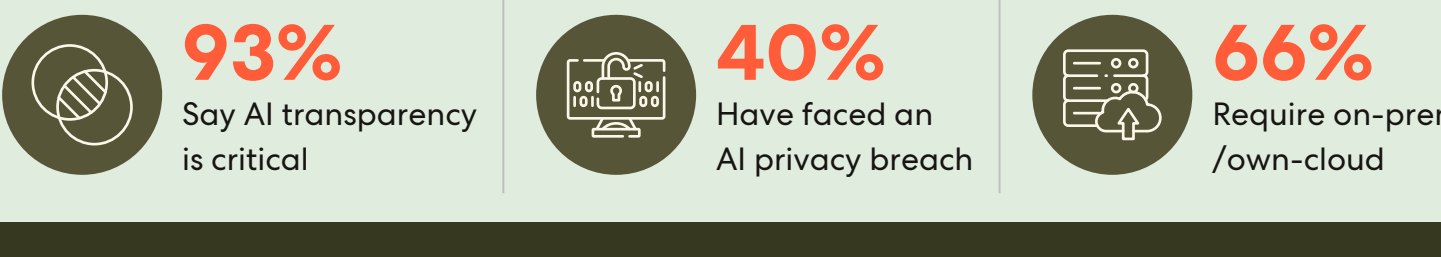
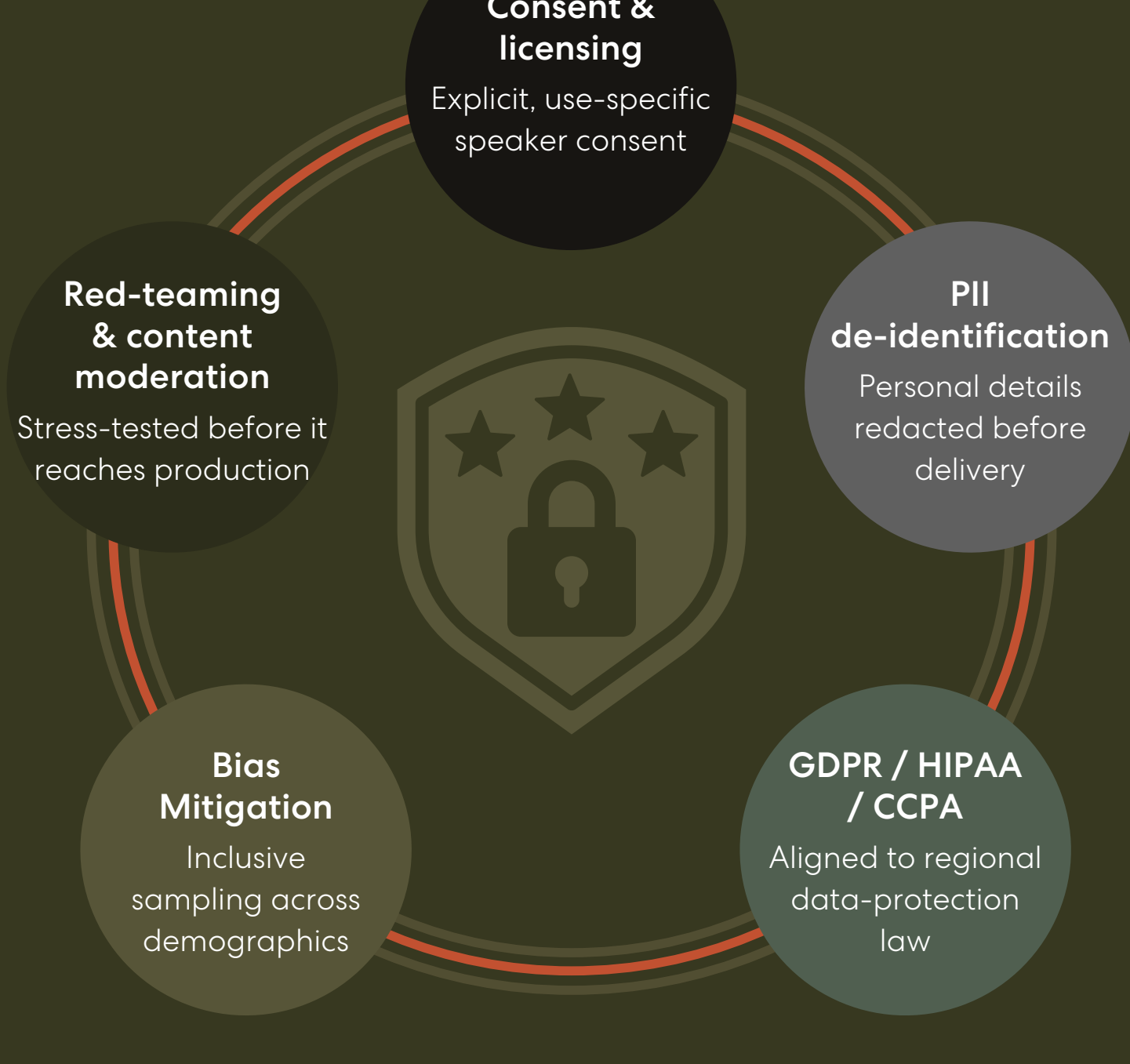
Volume alone doesn't make a dataset useful - these are the checks that decide whether it will hold up in production.

- | | |
|--|--|
| Long-tail intent coverage
Rare requests represented, not just the top 10 | Inter-annotator agreement
Consistent labels across multiple annotators |
| Error-rate tiers
Correct / minor / significant, tracked separately | Bias & drift monitoring
Checking for demographic & topic gaps |
| Benchmark validation
Measured against known standards before delivery | |

PRIVACY, ETHICS & COMPLIANCE

Trust is Built into the Data Layer

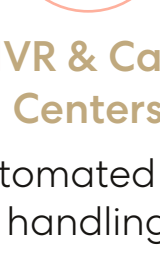
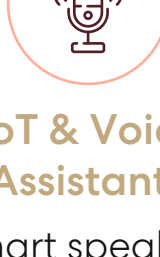
Speech data carries a person's voice & words - governance has to be designed in from collection onward, not patched on before shipping.



APPLICATIONS

Where Conversational AI Shows Up

Every one of these runs on the same underlying pipeline - the difference is the domain vocabulary, compliance needs, and voice data each is trained on.

- | | | | |
|---|---|---|--|
|  Customer Support
FAQ & service bots |  IVR & Call Centers
Automated call handling |  Banking & Fraud
KYC & fraud alerts |  Healthcare
Scheduling & triage |
|  HR & Onboarding
Policy Q&A, onboarding |  IoT & Voice Assistants
Smart speakers & wearables |  Retail & Commerce
Conversational shopping |  Accessibility & Translation
Multilingual, inclusive UX |

SHAIP EXPERTISE

Better Conversational AI Starts with Better Data

10+ years building the speech & language datasets behind production-grade conversational AI.



- Speech Data Collection
- Transcription
- Annotation
- De-Identification
- 100+ Languages

[Explore Shaip's Data Services →](#)