

WHY MULTILINGUAL SPEECH DATA IS CRITICAL FOR GLOBAL AI

Your voice model won't scale unless your dataset speaks the world's languages.

MOST VOICE AI MODELS FAIL IN REAL-WORLD SETTINGS

Key Challenges

1. Accent mismatch - inaccurate results
2. 40 - 60% accuracy drop in low-resource languages
3. Dialect and tone variations break recognition
4. Background noise disrupts ASR performance
5. Limited demographic diversity creates bias

STAT: 1 in 3 speech models fail due to linguistic diversity gaps.

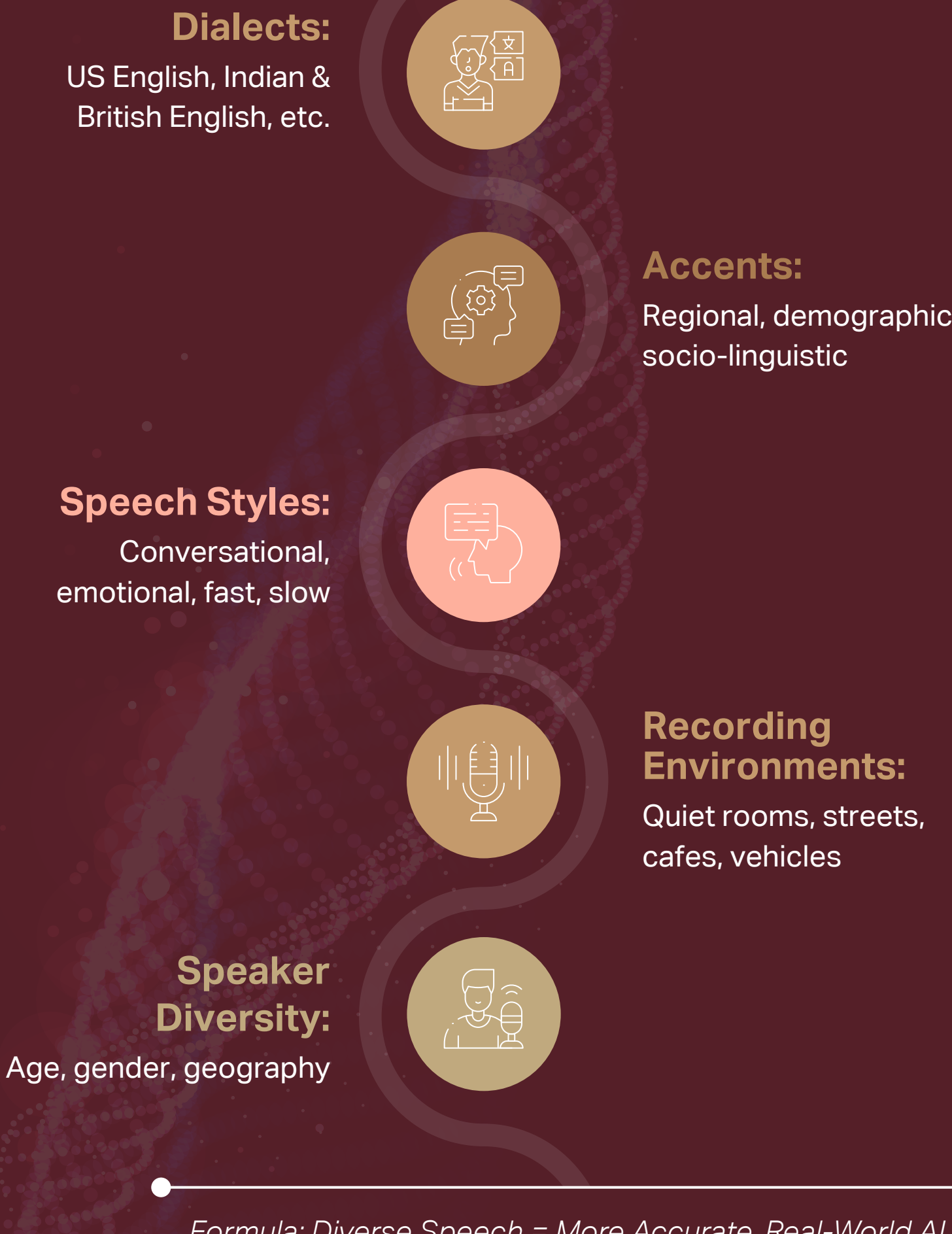
AI MUST UNDERSTAND REAL PEOPLE IN REAL CONDITIONS

Why Multilingual Matters

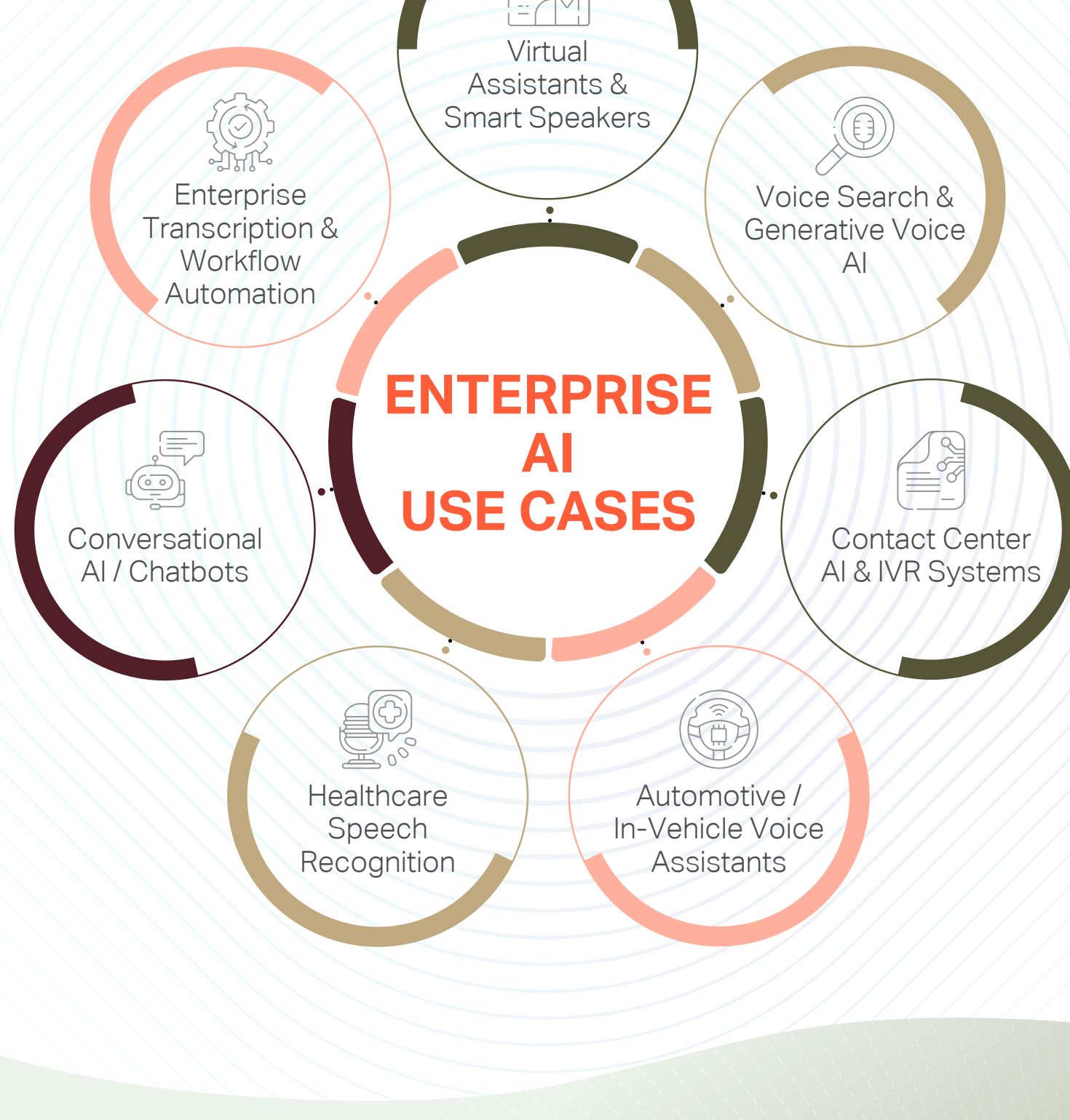


MULTILINGUAL SPEECH DATASET DNA

A Powerful Speech Dataset Covers



Formula: Diverse Speech = More Accurate, Real-World AI



HOW SHAIPI ENABLES WORLD-CLASS SPEECH AI

- » Native speakers across 60+ languages
- » Real-world + studio-quality recordings
- » Large-scale multilingual data collection
- » Speech segmentation, transcription and labeling
- » Multi-layer QA workflow
- » GDPR, HIPAA & enterprise-grade compliance
- » Fast delivery for global AI teams

RESULT: High-performing, inclusive, scalable speech models.

BUILD VOICE AI THAT UNDERSTANDS THE WORLD

Follow Shaip for more insights on Speech AI, ASR, & Multilingual Datasets