



M!AGiC DATA
Data Set Your Mind

CONTENTS

01

Company Overview

02

Product And Service

03

Data Security Compliance

01 Company Overview

COMPANY OVERVIEW



ASR

NLP

TTS

Core Service

MD Datasets

Annotation Platform

Data Collection & Annotation

One Stop Data Solution for Human-Machine Interaction



Data Solution for AI Industries



Customer Services

- Retailer Customer Service
- Finance Customer Service
- Telecom Customer Service



IOT

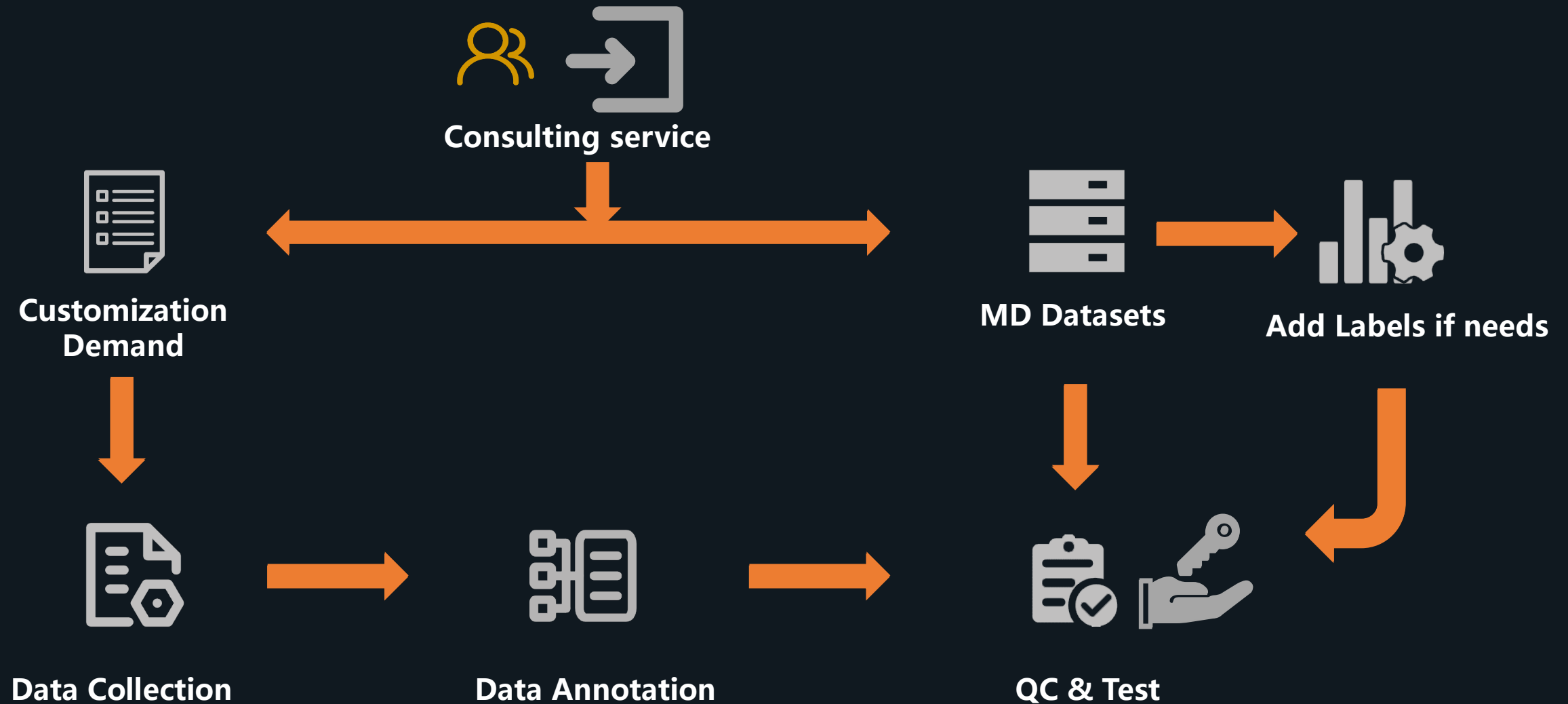
- Mobile Assistant
- Smart Home Assistant
- Smart Car Assistant
- Chatbot Conversation



Social Media

- Smart Subtitles
- Content Extraction
- Content Audit
- Content Relevance

Data Solution Workflow



COMPANY ADVANTAGE

Magic Data
was established.
Nuance Group Vendor

Microsoft
Master Vendor

TikTok Vendor



Tencent
Group Vendor

Qualcomm
vendor

Intel
vendor

ABOUT MAGIC DATA



Dr. Qingqing Zhang

Founder & CEO



Acoustic researcher, Data Scientist, and Conversational AI Pioneer

Former Associate Researcher at IACAS

Postdoctoral researcher at CNRS-LIMSI

Fortune “The Most Powerful Women 2021”

CYZone “Top Female Founder 2021”

Member of Committee of Acoustics/Automobile/Female Worker of CCF

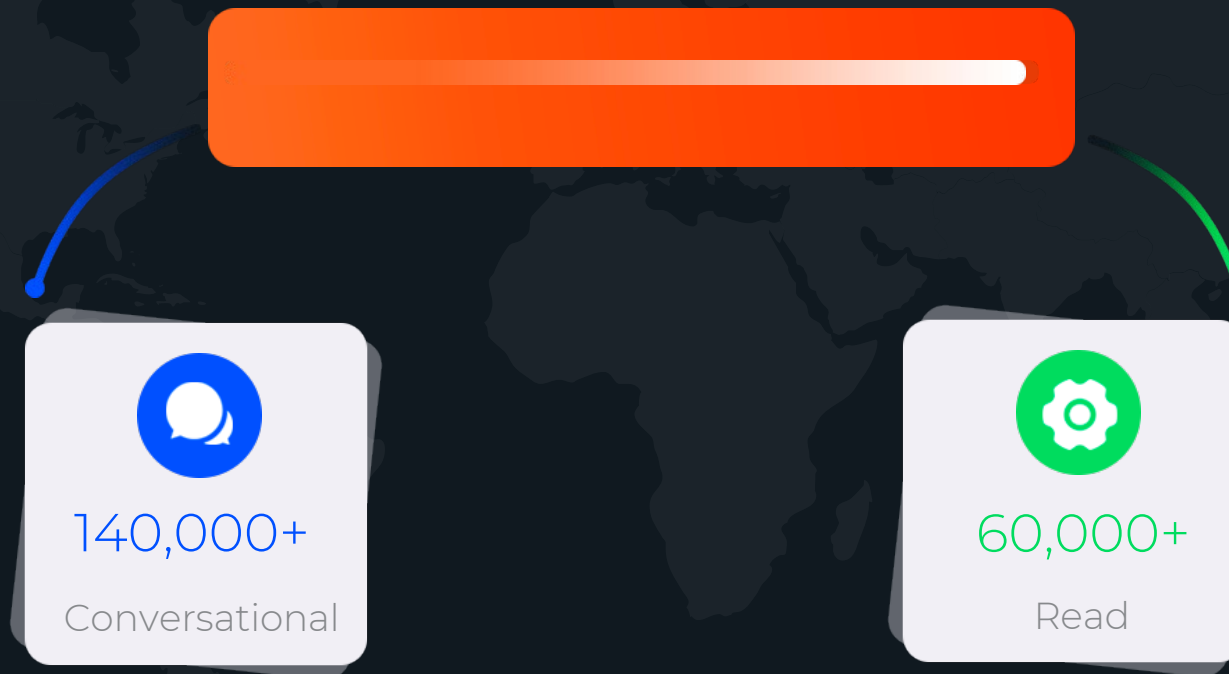
CAS Outstanding Scientific and Technological Achievement Award

Microsoft Research Asia Fellowship Award

02 Product and Service

MD Datasets

200,000 H +



One of the Most Comprehensive **Conversational** Dataset Providers

MD Datasets



CHINESE

MANDARIN
SICHUAN DIALECT
SHANGHAI DIALECT
MINNAN DIALECT
CANTONESE
etc



ENGLISH

AMERICAN ENGLISH
AUSTRILIAN ENGLISH
MALAY ENGLISH
THAI ENGLISH
TURKISH ENGLISH
HONGKONG ENGLISH
CHINESE ENGLISH
SINGAPORE ENGLISH
etc



EUROPEAN

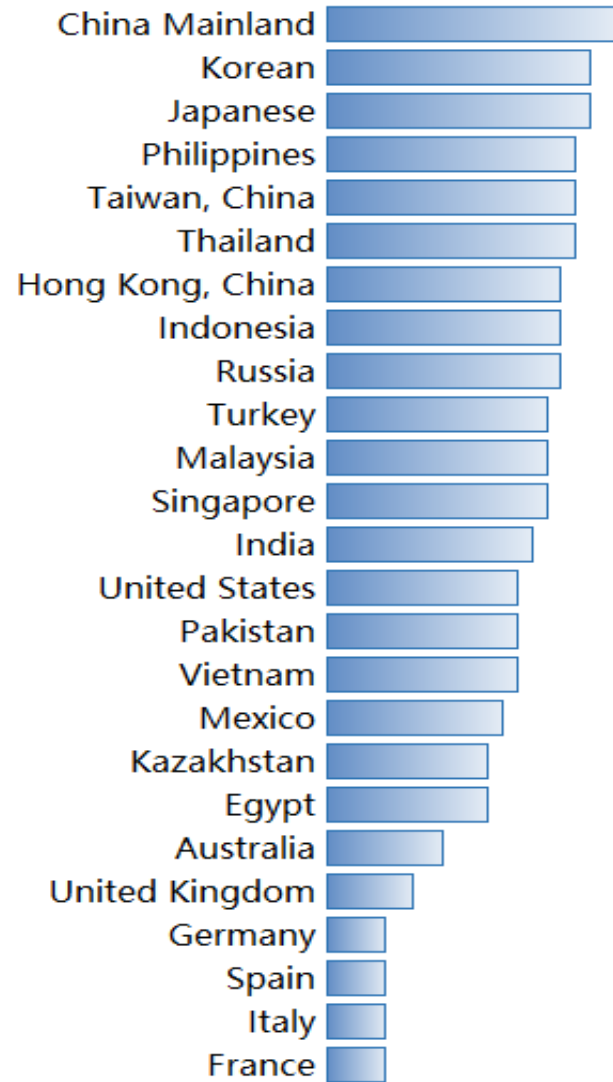
FRENCH
ES-SPANISH
GERMAN
ITALIAN
etc



OTHERS

KOREAN
JAPANESE
THAI
INDONESIAN
TAGALOG
MALAY
RUSSIAN
etc

MD Datasets



1

• Overwhelming competitive conversational resources

2

• Rich resources over all Asian Languages

- 60+ languages

- 80k+ hours

Conversational speech in
Chinese

- 60k+ hours

Conversational speech in
foreign languages

ASR Dataset

Dataset Name	H
Peninsular Arabic Conversational Speech Corpus	500
German Conversational Speech Corpus	500
Russian Speech Corpus	500
French Conversational Speech Corpus	200
Filipino/Tagalog Conversational Speech Corpus	500
Korean Conversational Speech Corpus	3000
Malay Conversational Speech Corpus	1000
Brazilian Portuguese Conversational Speech Corpus	200
Japanese Conversational Speech Corpus	3000

Chinese Example

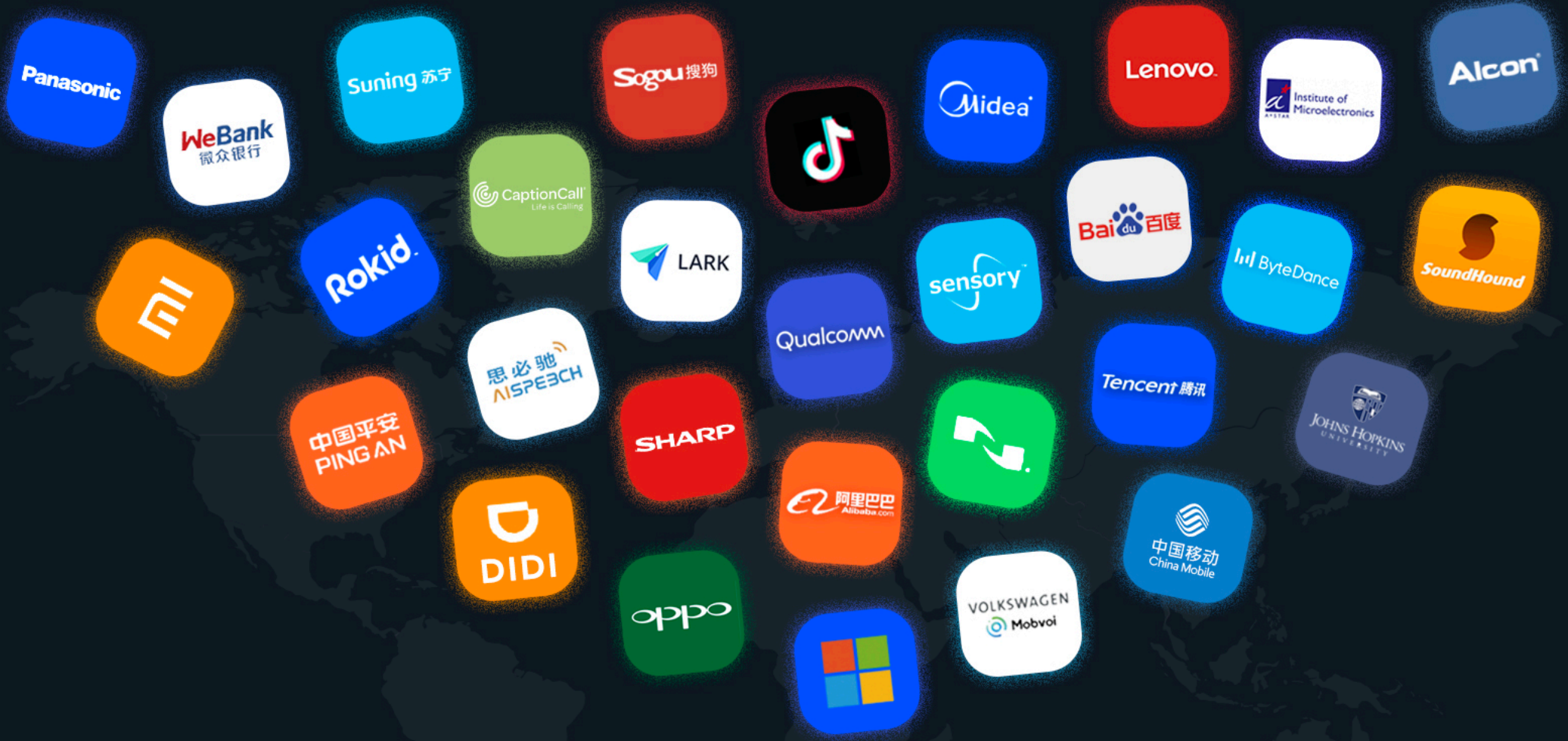
Dataset Name	H
Chinese Conversational Speech Corpus	4000+
Sichuan Conversational Speech Corpus	2000
Cantonese Conversational Speech Corpus	2000
etc	



Dataset Name	H
Thai Speech Corpus	500
Turkish Conversational Speech Corpus	1000
Peninsular Spanish Conversational Speech Corpus	200
Italian Conversational Speech Corpus	500
Hindi Speech Corpus	500
Bahasa Indonesian Conversational Speech Corpus	1000
English Conversational Speech Corpus	2000
Chinese Conversational Speech Corpus	5000
Chinese-English Code-Mixing Speech Corpus	3000

English Example

Dataset Name	H
English Conversational Telephone Speech Corpus	1000
Japanese English Conversational Speech Corpus	200
Singaporean English Speech Corpus	400
etc	



OUR CLIENTS

Contact us: open@magicdatatech.com

SUCCESS STORY



“For the conversational speech recognition in noisy environment, the nearly 10,000 hours of Magic Data conversational speech dataset **brought significant improvement to our model performance.**

The error rate of speech recognition in natural conversation is reduced by approximately **30%.**

Besides, we are also pleasantly impressed by error rate reduction for command and control speech recognition within noisy environment, which is surprisingly **10%.** It shows that the speaking randomness of the data can not only help the model to enhance the ability to recognize the natural speaking, but also strengthen the robustness of the model performance against noise.”

Former Nuance Greater China R&D Director - Dr. Xu Ran

SUCCESS STORY



AI requires training data, but collecting and labeling data is not always the best use of AI experts.

With Magic Data handling your speech data needs, you can focus on algorithm development.

--- Principal Scientist Advisor Daniel Povey

03 Data Security Compliance

DATA SECURITY COMPLIANCE

Consulting a worldwide accounting firm to improve our data security measures and protocols

Engaged a worldwide law firm (Kennedys Law LLP) to draft various documentations to compliant with GDPR and ISO/IEC 27001 & ISO/IEC 27701:2019



Conduct annual supplier security assurance audit to meet a worldwide customer's vendor selection requirement

Regular staff training on data security compliance are conducted by senior member of Magic Data and attorneys within the past 12 months

GDPR Compliance Documentations Which Is Used In Both Our Online APP And Offline Projects

Data Protection Documentations

A

APP software license agreements reviewed by Kennedys Law LLP

B

Consent letter, Privacy Notice and Task Participation Terms reviewed by Kennedys Law LLP

C

Company policy on Personal Data and Privacy Protection

D

Response Mechanism and Contingency Plan for Data Breach Incidents



Thank You

DATA SET YOUR MIND

www.magicdatatech.com

Contact us: open@magicdatatech.com