

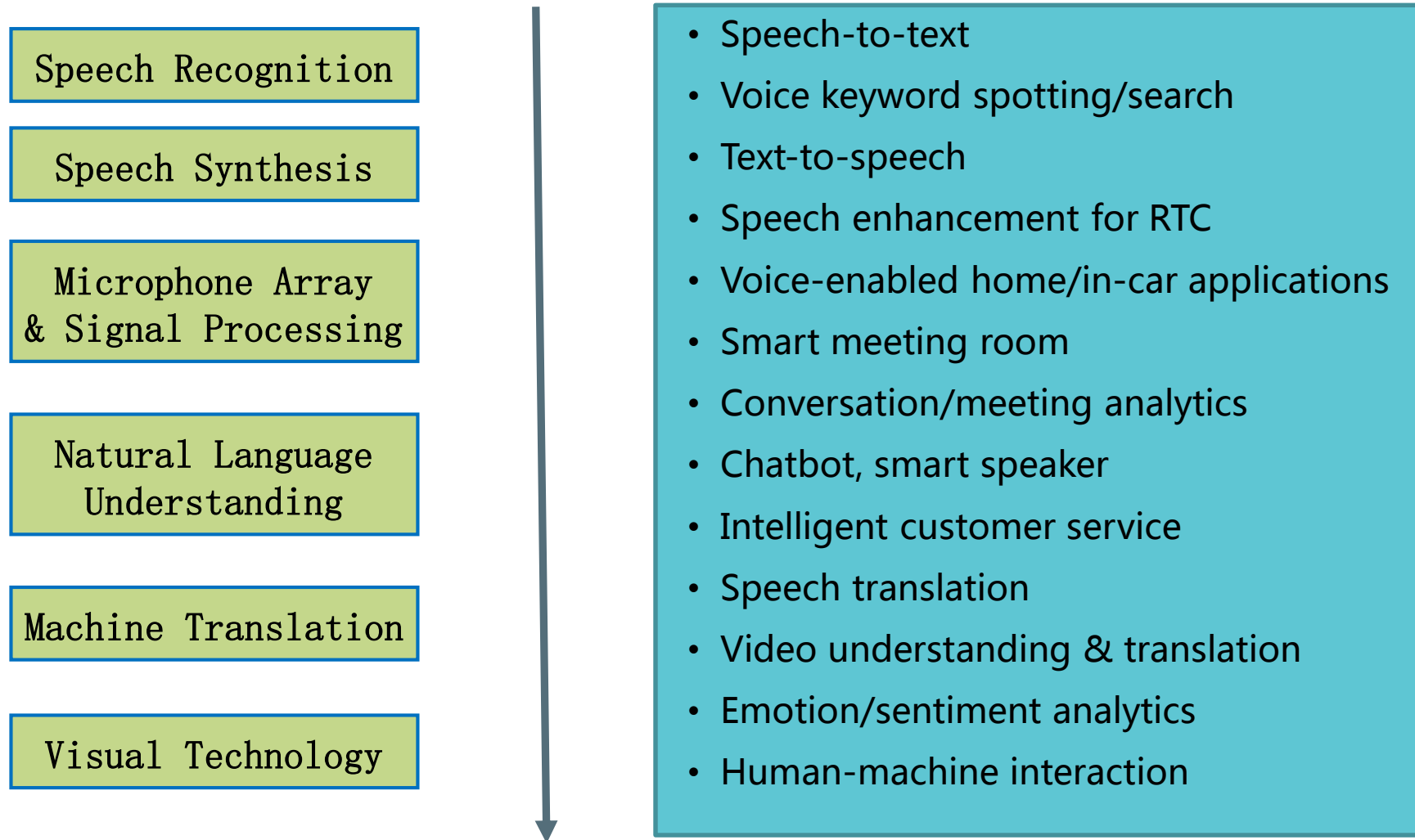


Unlocking The Potential of AI: from Machine Intelligence Perspective

Bin Ma

Singapore R&D Center, Alibaba DAMO Academy

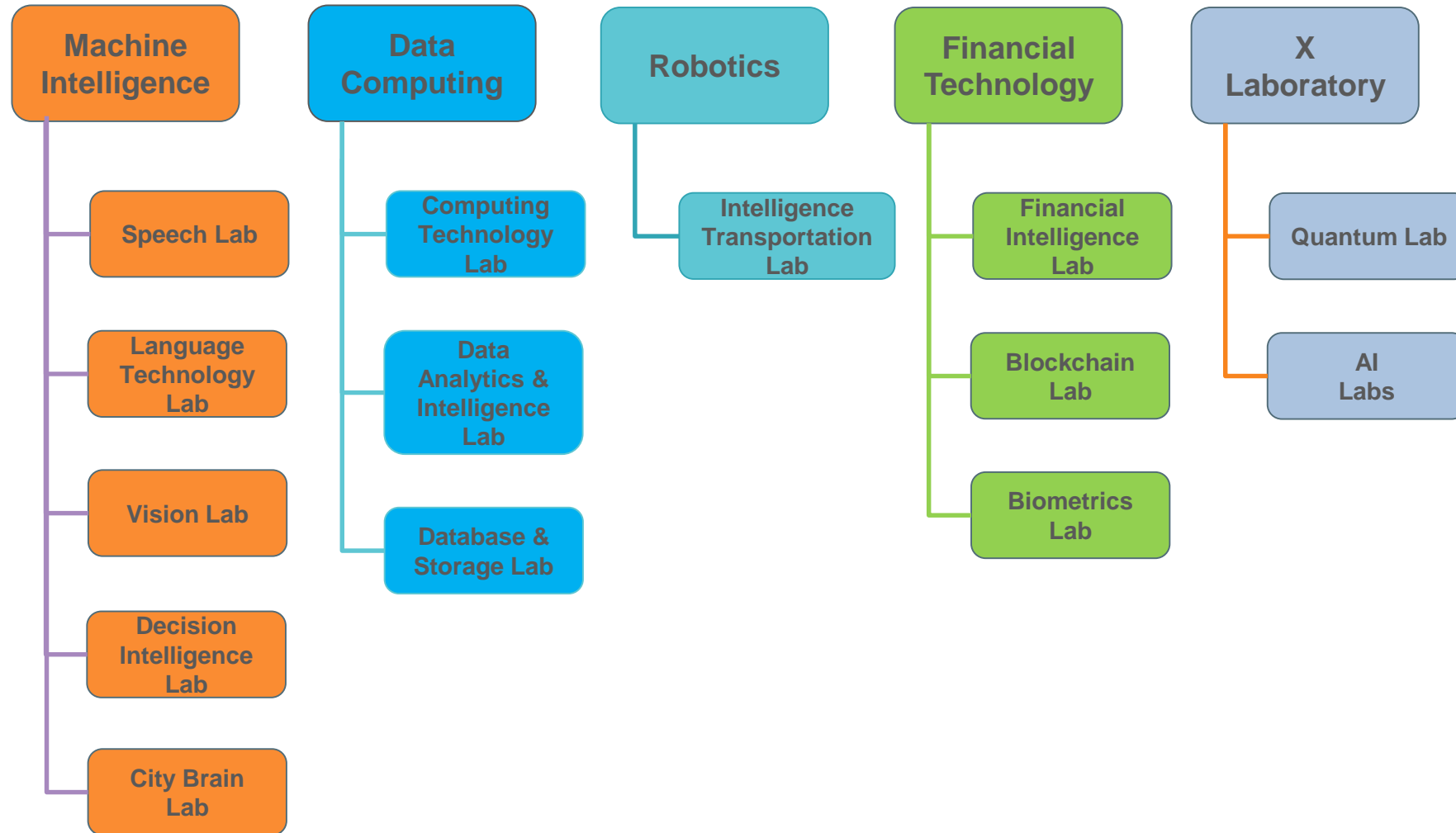
Machine Intelligence Technologies & Applications



Satisfied and Robust Performance is the KEY

Alibaba DAMO Academy

5 Research Areas | 14 Laboratories



Machine Intelligence Division (MinD)

7 Research Centers
in China, US, Israel and Singapore

1000+ Researcher & Engineers
45% with PhD degree

139

Journal Articles

200+

Patents Filed

28

International Academic Institutions
(Fellow, Member)

21

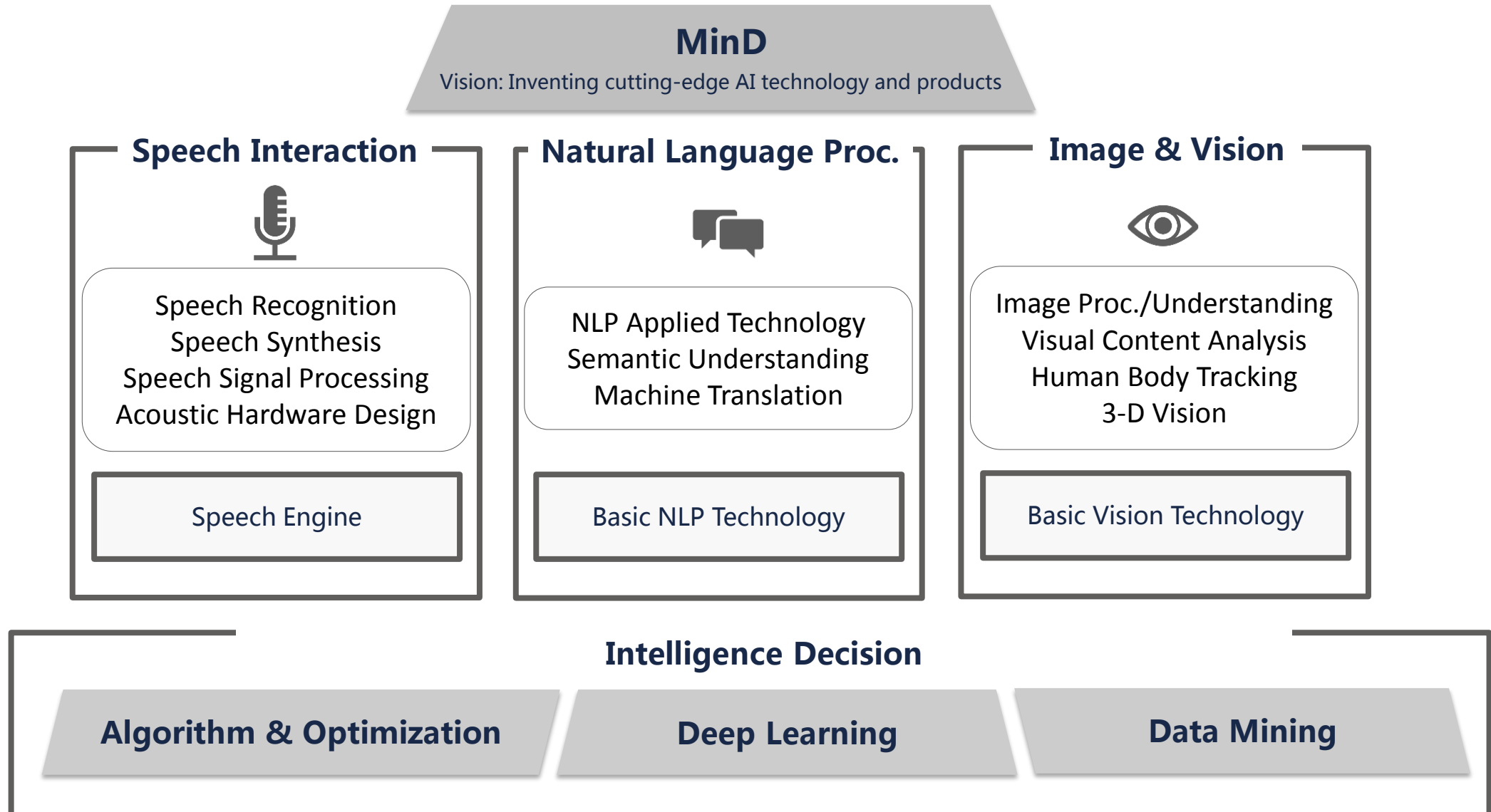
International Benchmark Competitions
(1st place)

3

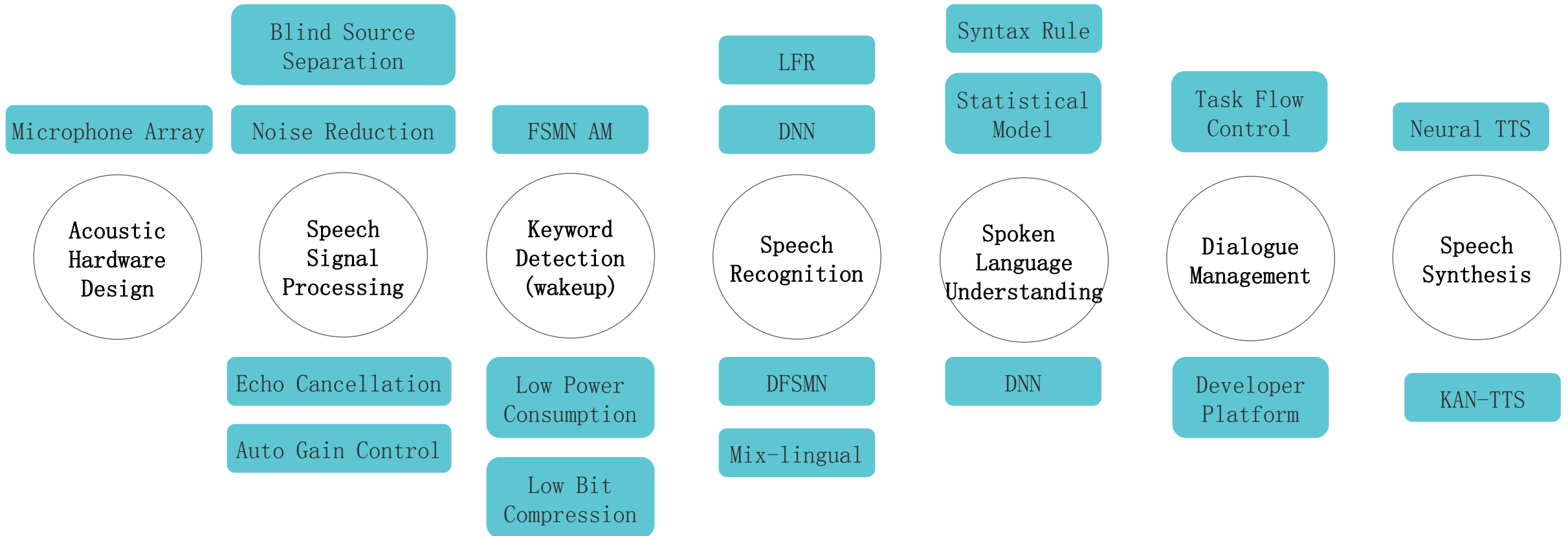
National/Provincial Awards

from 2017 to 2019

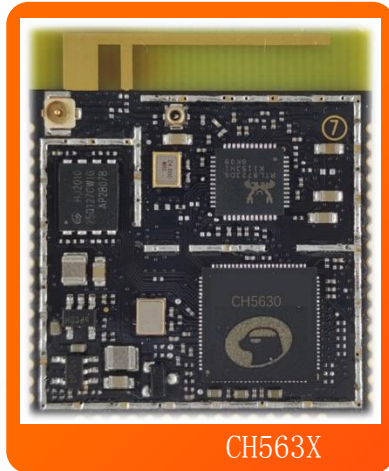
Machine Intelligence Division (MinD)



Speech Interaction for Human-Machine Communication



IoT Voice Module for Speech Interaction



High Performance

2-4mic with highly anti-noise front-end algorithms



Inexpensive

Multi-core heterogeneous chip with RTOS system



Low Power Consumption

Working current < 150mA
Stand-by current < 10mA

IoT Voice Module for Speech Recognition	Blind Source Separation based MIMO	<ul style="list-style-type: none">• Environment noises, room reverberation and device echo are modeled as independent signals, and are separated under a unified BSS-MIMO framework that maximizes signal independency.• The separated multi-stream outputs represent different signal processes and are fed to subsequent modules, e.g. ASR.
	Signal Processing + Keyword Spotting	<ul style="list-style-type: none">• Joint optimization of KWS and BSS in the training.• Use KWS feedback to enhance target speech.

Speech Acoustics and Signal Processing



Microphone Development Board
(Polar Array)

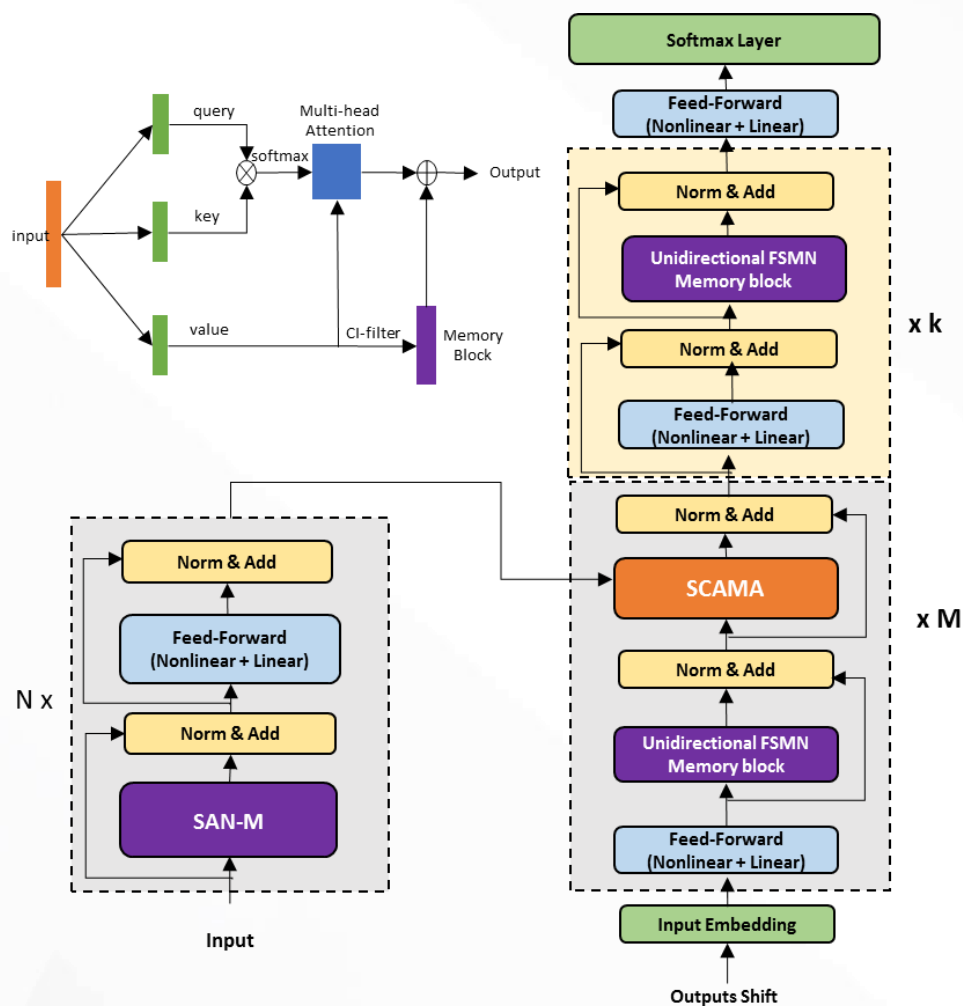
New Future on Cloud



All-in-One for Audio-Visual Meeting
(Linear Array)

Technology	Differential Directional Microphone Array	<ul style="list-style-type: none">• High directional sound localization• Distance voice acquisition (more than 5m)• Strong de-reverberation capability (supporting RT60 of 1.5 ~ 2s)• High quality voice (improving SNR by 20dB for low frequency voice)
	AGC, ANS, AEC	<ul style="list-style-type: none">• State-of-the-art audio 3A algorithms• Full audio spectrum (100 ~ 20kHz) coverage• Elimination of 99% playback echo and less than 3% distortion of human voice by improving the transparency of double-talk

Speech Recognition (ASR) Technology



- DFSMN

- Feedforward Sequential Memory Networks (FSMN)

- “Deep-FSMN for Large Vocabulary Continuous Speech Recognition”, ICASSP 2018

- Open source at <https://github.com/alibaba/Alibaba-MIT-Speech>

- End-to-end

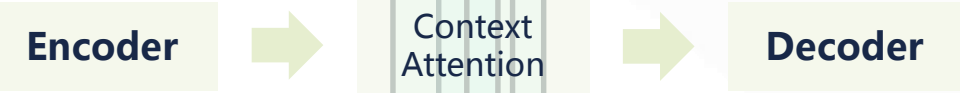
- Integrated modeling of E2E-ASR and post-processing for a small-footprint, high performance, on-device ASR system

- Memory equipped self-attention network for efficiently combining DFSMN and self-attention for both local and global modeling, “Memory Equipped Self-Attention for End-to-End Speech Recognition”, INTERSPEECH 2020

- Streaming chunk-aware multi-head attention for constructing streaming end-to-end ASR with a single model to support multiple latency requirements, “Streaming Chunk-Aware Multihead Attention for Online End-to-End Speech Recognition”, INTERSPEECH 2020

Natural Language Understanding

Multilingual Neural Machine Translation



Machine Reading Comprehension



Top performance at the WMT18 on English->Chinese, English<->Russian, English<->Turkish

		output language									
input language	Czech	popel CUNI-Trans									
	German	jschamper RWTH									
	English	benjamin.marie NICT	benjamin.marie NICT	Alibaba-MT Alibaba-en	Alibaba-MT Alibaba-en	Alibaba-MT Alibaba-en					
	Estonian	tilde tilde-nu-n									
	Finnish	benjamin.marie NICT									
	Russian	Alibaba MT Alibaba	Alibaba-MT transforme								
	Turkish	MF Translator Tencent									
	Chinese										

1st place in 2018 SQuAD reading comprehension competition, with Exact Match score of 82.44, which beats out the human high of 82.304

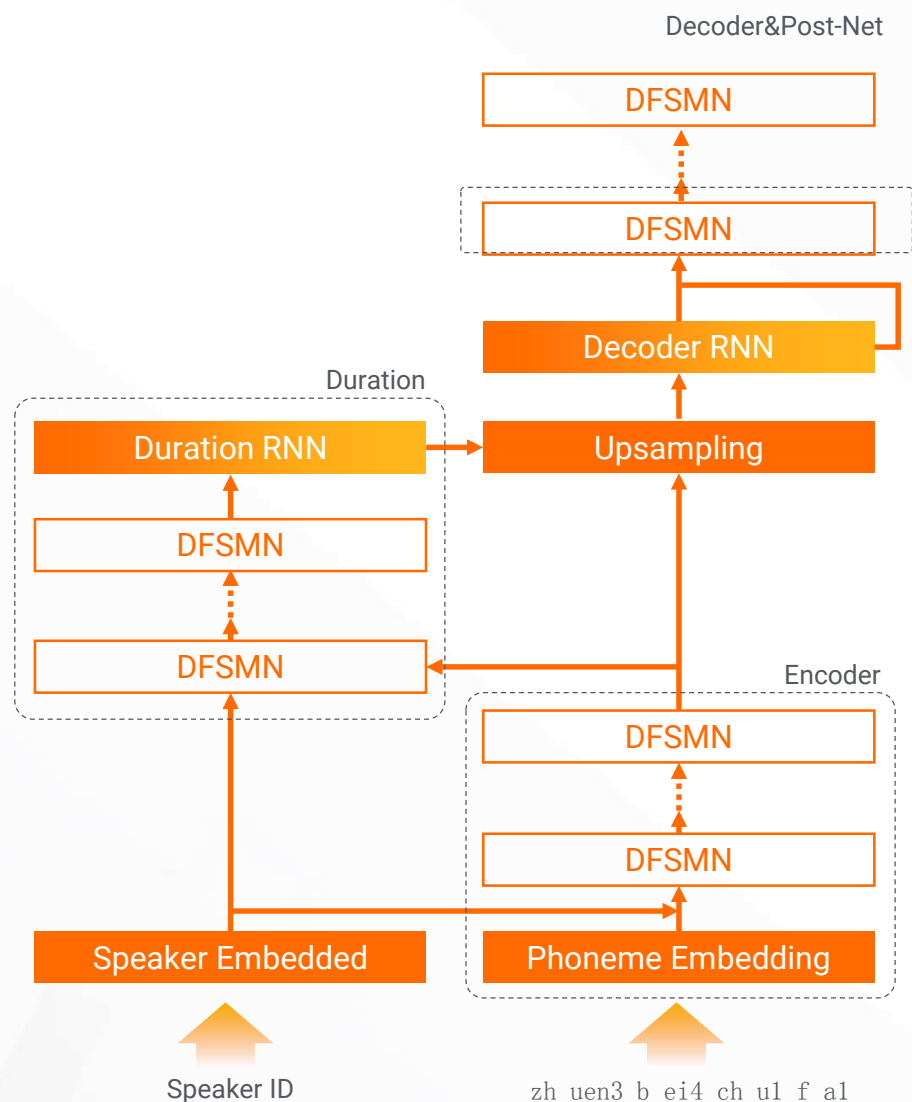
THE WALL STREET JOURNAL
Subscribe | Sign In

TO JOURNAL

The Morning Download: Alibaba, Microsoft AI Beats Human Reading Comprehension

MIT Technology Review
Artificial Intelligence JUL 9
Alibaba has claimed a new record in AI language understanding

Speech Synthesis (TTS) Technology



- KAN-TTS

- Knowledge-aware Neural (KAN) TTS with both traditional and end-to-end TTS techniques, dramatically improving TTS quality
- Customized voice for a user with the voice recording of only 10 mins

- Lite KAN-TTS

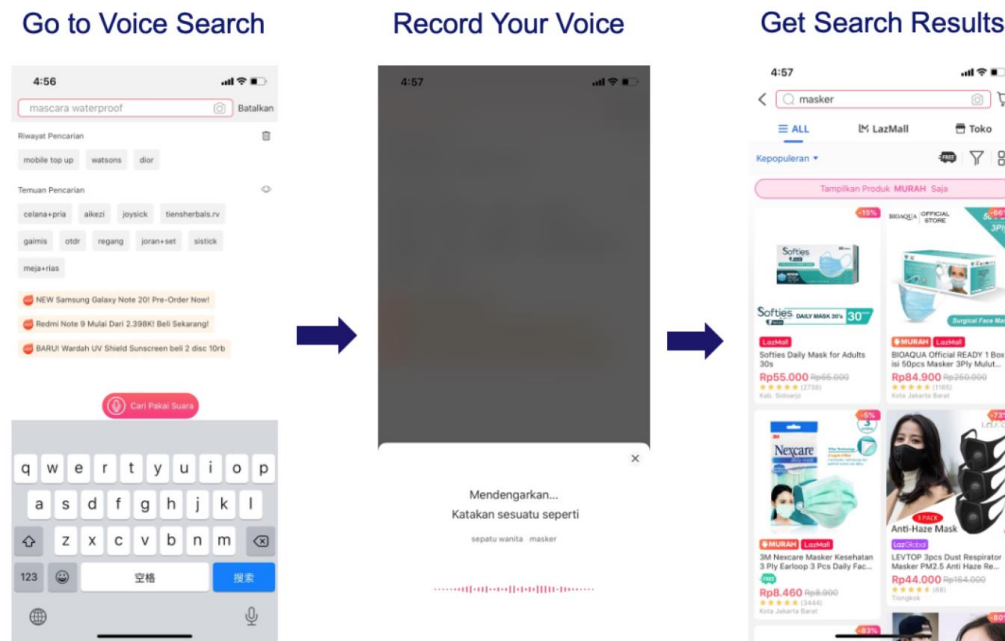
- Acoustic model calculation reduced by 100 times
- Acoustic model size compressed by 35 times
- No significant loss in performance

- Mix-lingual TTS

- Cross-lingual voice conversion
- Mix-lingual TTS with mono-lingual speaker's voice, [“Towards Natural Bilingual and Code-Switched Speech Synthesis Based on Mix of Monolingual Recordings and Cross-Lingual Voice Conversion”](#), INTERSPEECH 2020

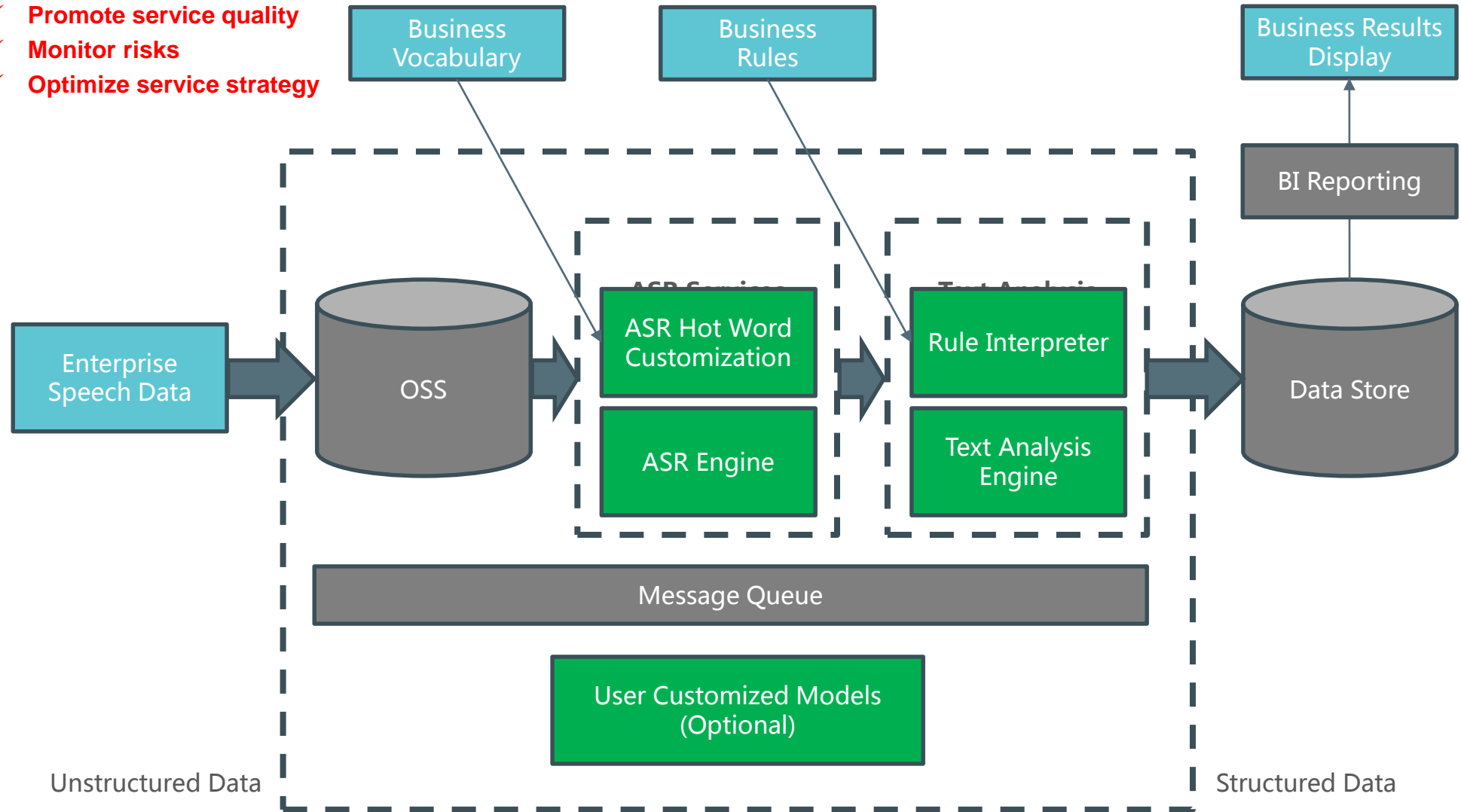
Voice Search for e-Commerce

- Quick and natural user interaction.
- Voice search function for all LAZADA users in Indonesia, Singapore, Malaysia, Philippines, Thailand and Vietnam.
- LAZADA is the first e-commerce platform to offer voice search experience in Southeast Asia.



Intelligent Quality Control for Call Centers

- ✓ Promote service quality
- ✓ Monitor risks
- ✓ Optimize service strategy



Intelligent Customer Service for Alibaba's Single's Day (11.11)

Double Eleven is the annual online promotion day organized by Alibaba Tmall since November 11, 2009.

Alipay 95188 hotline has been upgraded with speech dialogue technology.

- 98% of the telephone customer services were completed by AI voice customer service
- 97% of online customer services were completed by AI voice customer service, equivalent to a workload of 85,000 manual customer service juniors



2019 Double Eleven (11.11) created a new highest record in China

Alipay 95188 Call Center

Multi-Modal Speech Interaction

“Look” + “Listen”

- Multi-modal man-machine interactive technology for integrating speech and visual information together.
- Capture users' voice accurately at noisy conditions.



UHD Periscopic Camera



- Wake-up free for a natural man-machine interaction
- High ASR accuracy at strong noisy environments

8-Mic Large-scale Array



Voice-activated Ticket Machine (Shanghai Metro)



- Integrated with AMAP & Alipay for auto routine plan and online payment
- Deployed in Shanghai, Beijing, Chengdu and trials in other 8 Cities

Information Kiosk with Intelligent Speech Interaction

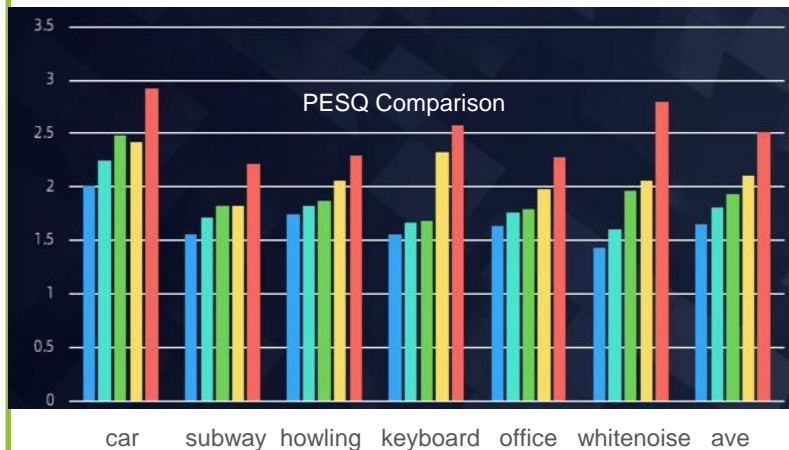
- User friendly AI platform for information query
- Integrated technologies of multi-modal wake-up, speech recognition, natural language understanding, dialogue management, speech synthesis, avatar, user experience design



Audio-Visual Meeting Solutions

Signal Processing & Enhancement

- Conventional signal processing and deep learning approaches for 3A algorithms (AEC, ANS, AGC)
- Speech enhancement for reducing de-reverberation, howling removal, bandwidth expansion and automatic voice quality assessment



Audio Hardware

- Advanced acoustic design and audio algorithm for all-in-one unit for audio-visual meeting solution

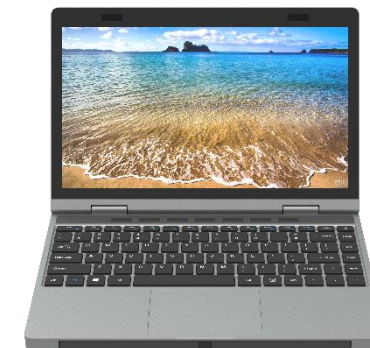


- Microphone hardware based on differential directional microphone array technology for high quality of audio signals, speaker separation and speech recognition



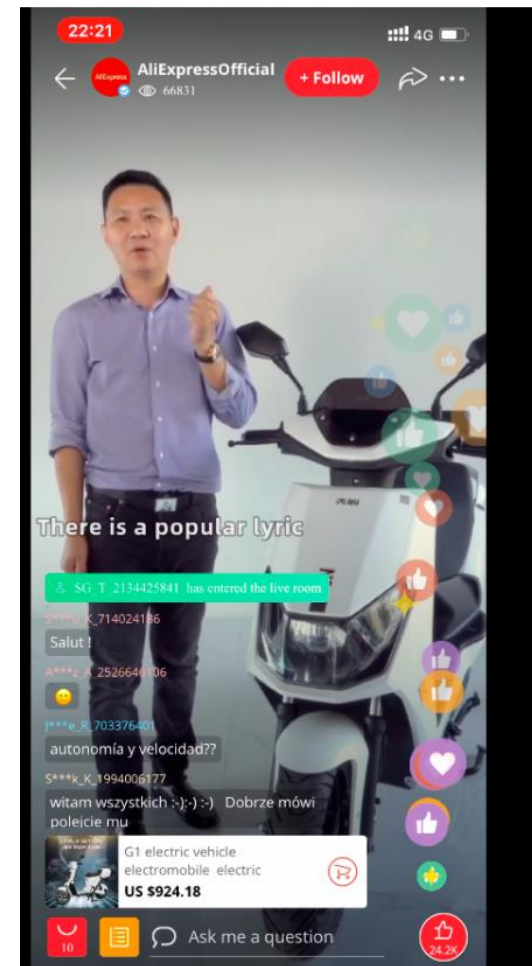
Meeting Transcription/Summary

- Speech to text transcription function for DingTalk and AliLang meeting recordings
- Integrated solution of hardware/software for a number of application scenarios like meetings and interviews

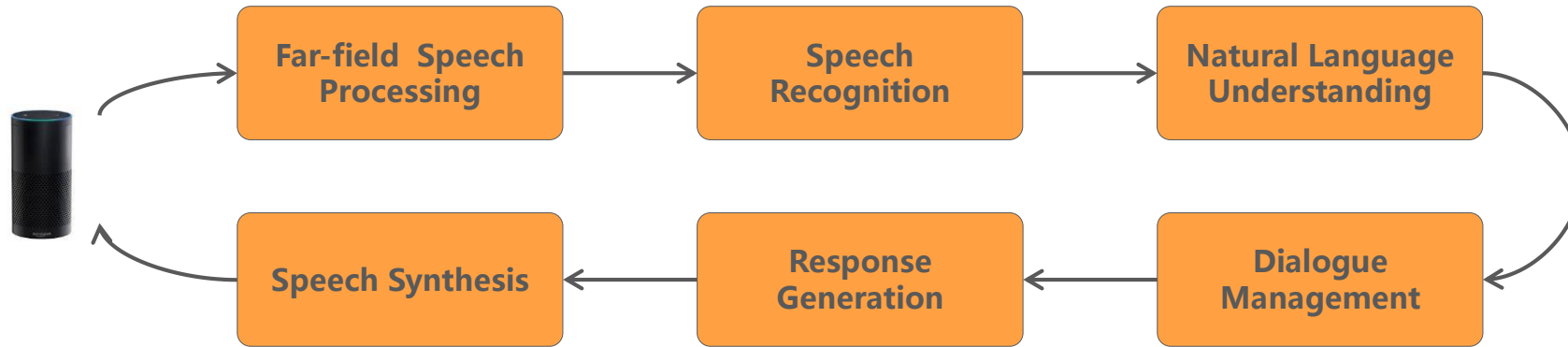


Speech Translation

- Speech recognition + machine translation, breaking the language barriers
- Communicative translation for meetings
- Simultaneous translation of speeches
- Simultaneous translation of live broadcast
 - In keeping with the rising trend of livestreaming in e-commerce, AliExpress, a global retail marketplace under the Alibaba Group, unveiled the world's first real-time livestreaming translation feature, from Chinese to English, Russian and Spanish, on an e-commerce platform, to tackle the biggest obstacles in cross-border e-commerce for SMEs.



Smart Speaker



天猫精灵
TmallGenie



Home Automation



Music



Nursery Rhyme



Audio Book



Voice Shopping



Weather Forecast



Takeout Order



Entertainment News



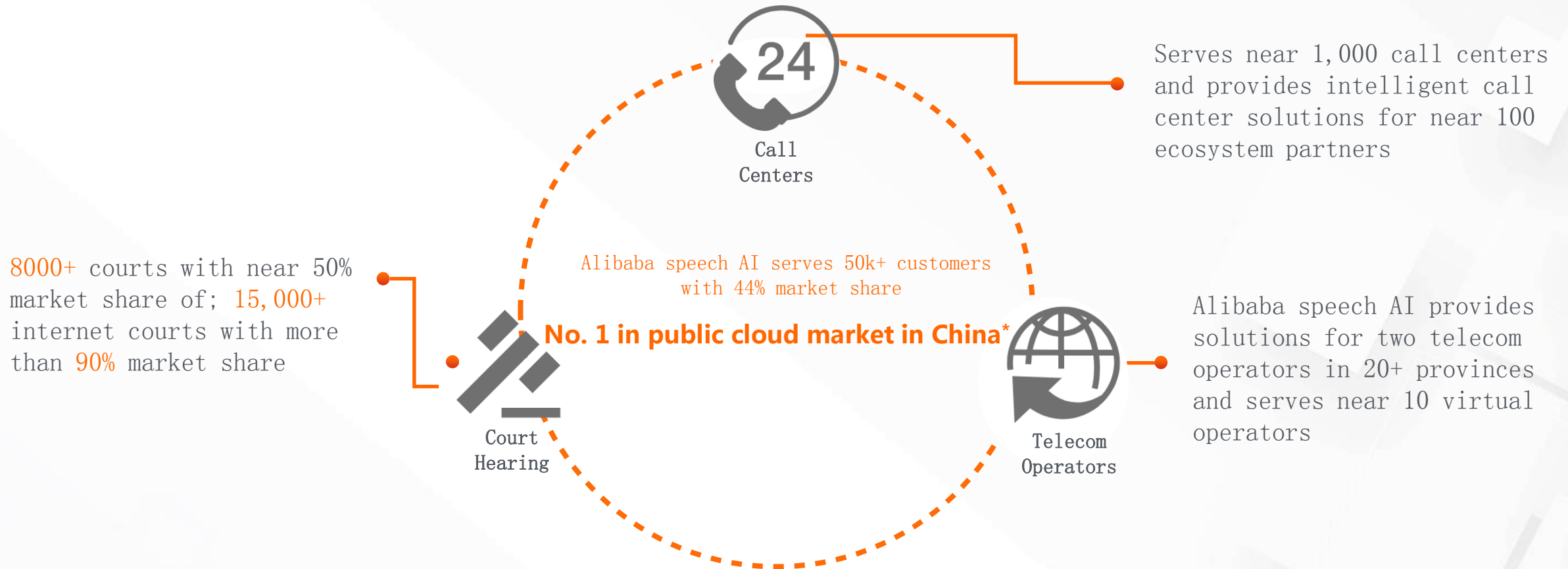
Mobile Phone Top-up



Calendar Reminder

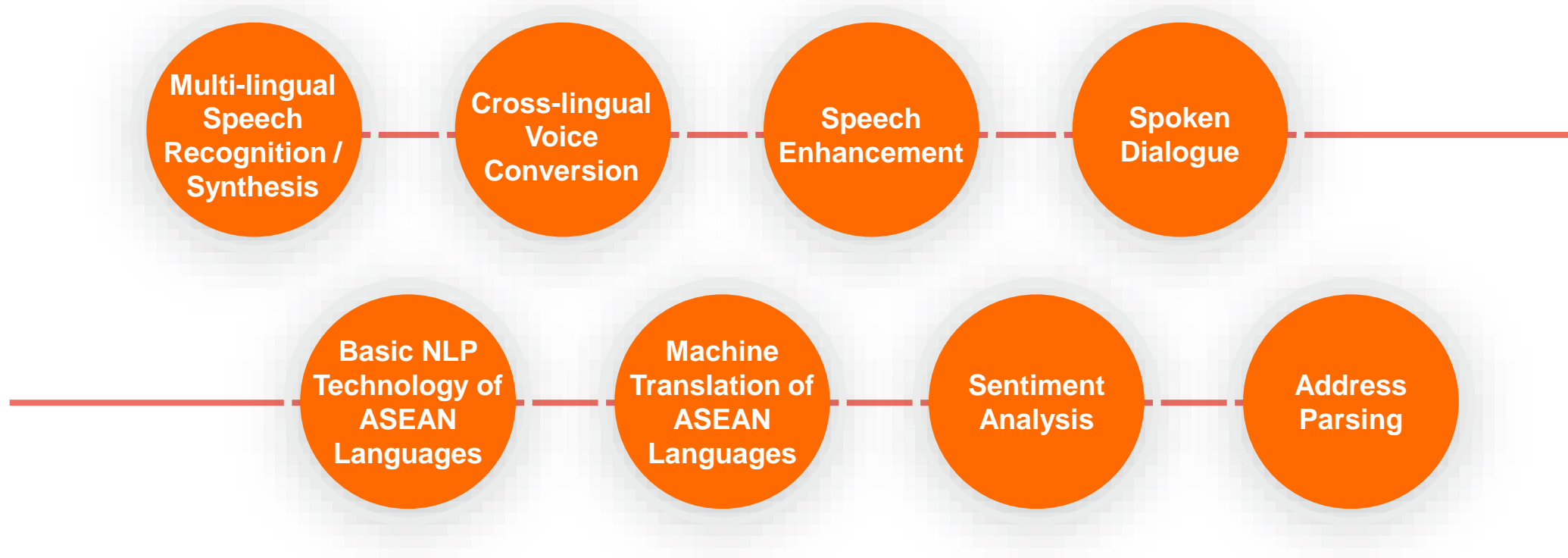
The top-selling smart speaker in China for three consecutive years from 2017 to 2019

Scale Ecosystem of Alibaba Customers with Intelligent Speech Technology



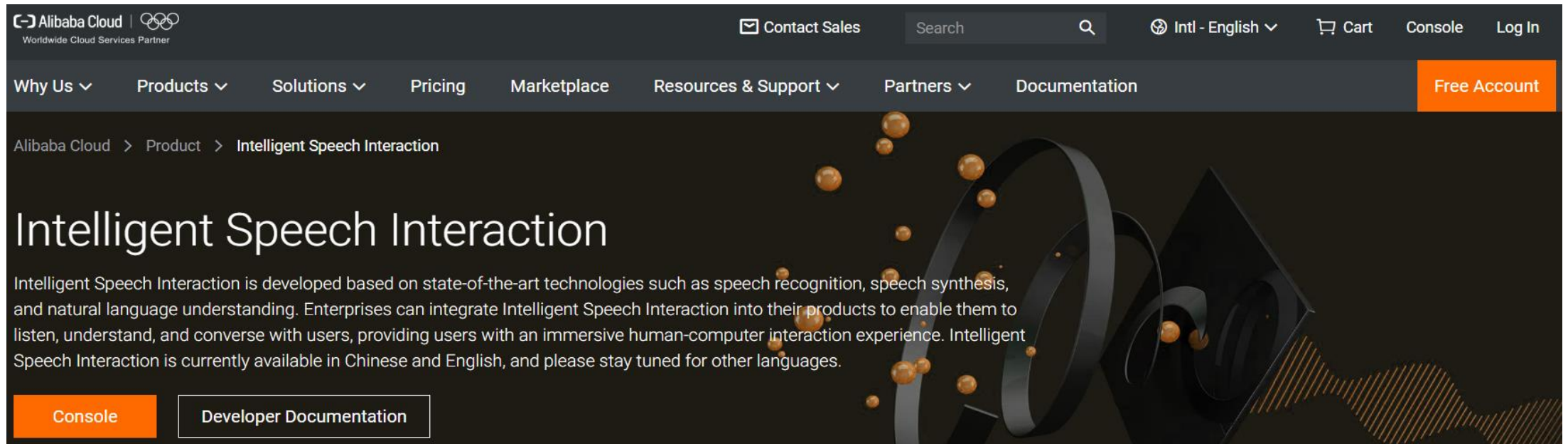
* Data sources from IDC 《中国人工智能云服务市场研究报告(2019)》
For reference purposes only. It does not necessarily reflect the views of Alibaba.

Singapore R&D Center



Multi-lingual Speech Products in Singapore

- International station in Singapore has been launched for intelligent speech products
- ASR for English, Chinese (Mandarin), Indonesian, Vietnamese, Thai, Cantonese, Japanese, Korean, Russian, Spanish, Arabic have been released, and ASR for another 7 languages will be released by Dec 2020
- TTS for English, Chinese (Mandarin), Cantonese, Japanese have been released, and TTS for Indonesian will be released by Dec 2020
- Self-learning modeling platform for providing an effective way for optimizing ASR performance by customers
- Quickly and flexibly scale to support various business concurrent requirements
- Data stored in Singapore



The screenshot shows the Alibaba Cloud website interface for the Intelligent Speech Interaction product. The top navigation bar includes the Alibaba Cloud logo, a search bar, and links for Contact Sales, Intl - English, Cart, Console, and Log In. Below this is a secondary navigation bar with links for Why Us, Products, Solutions, Pricing, Marketplace, Resources & Support, Partners, and Documentation, along with a prominent orange 'Free Account' button. The main content area features a dark background with a decorative graphic of orange spheres and a grey ribbon. The heading 'Intelligent Speech Interaction' is displayed in large white text. Below the heading, a paragraph describes the product's capabilities: 'Intelligent Speech Interaction is developed based on state-of-the-art technologies such as speech recognition, speech synthesis, and natural language understanding. Enterprises can integrate Intelligent Speech Interaction into their products to enable them to listen, understand, and converse with users, providing users with an immersive human-computer interaction experience. Intelligent Speech Interaction is currently available in Chinese and English, and please stay tuned for other languages.' At the bottom of the page, there are two buttons: 'Console' (orange) and 'Developer Documentation' (white with a grey border).

Alibaba-NTU Joint Research Institute

- In Feb 2018, Alibaba set up Alibaba-NTU Singapore Joint Research Institute, the first overseas research institute of Alibaba. The institute will focus on AI research and will involve a number of researchers from both organizations.

- 50 professors
- 100+ research fellow and students
- 10+ research fields
- Speech and NLP
- City Brain
- E-commerce
- Search and Recommendation
- Medical and Health
- Safety



Alibaba Talent Programme in Singapore

➤ Talent Programme

- In September, 2018, Alibaba launched “Alibaba Talent Programme” . The programme is supported by the Industrial Postgraduate Programme (IPP) of EDB, Singapore
- Collaboration with universities for local talent development

➤ Competitive Research Talent Selection

- Selected from 2000+ applicants (acceptance rate <2.5%) in the first three batches
- Outstanding academic ability (Lee Kuan Yew Gold Medal, Dean's list of NUS & NTU)

➤ Outstanding Research Capability

- All 6 submissions for EMNLP 2020 from IPP students were accepted
- 3 submissions for INTERSPEECH 2020 from one IPP student were accepted



Thank You !

