

51CTO WOTI



智能交互技术与物联网应用

梁家恩 博士
云知声董事长 / CTO

51CTO WOT

WOTD

World Of Tech
2017年12月1-2日

全球软件开发技术峰会

[深圳站]

报名咨询：010-68478816

议题提交：wot@51cto.com

市场合作：yangxh@51cto.com

商务合作：songjc@51cto.com

媒体合作：yankk@51cto.com

在线咨询（微信）：18401576051

团·购·享·受·更·多·优·惠

5折

优惠（截止8月31日）

现在报名，立省1400元/张

 **物联网时代到来：物联网 + 人工智能**

人工智能引领物联网时代到来

人工智能 (AI) 将改变交互形态 , 并充分发挥大数据价值。

AI 重定义 IoT : Internet of Things (物联网) => Intelligence of Things (智联网)

信息时代先驱

- 图灵机概念 (1936年)
- 诺伊曼架构 (1945年)
- 香农信息论 (1948年)

信息时代 (1969年后)

- 计算能力 (Moore's Law)
- 通讯能力 (Gilder's Law)
- 存储能力

移动互联 (2006年后)

- 移动互联 : 用户参与、大数据
- 云平台 : 存储 + 计算
- 终端应用 : 个性化
- 连接价值 (Metcalfe's Law)

智联网时代 (2010年后)

- 终端和数据规模指数增长
- IoT 交互和服务智能化
- 智慧生活 / 城市 / 制造
- 大数据价值 (? Law)

人工智能
概念提出

人工智能
第一次寒冬

人工智能
第二次寒冬

深蓝击败
Kasparov

Hinton 兴起
深度学习

IBM Watson
Apple Siri

百度 IDL
成立

微软 小冰
Amazon Echo

AlphaGo
击败李世石

1956年

1970年

1987年

1997年

2006年

2011年

2013年

2014年

2016年

云知声@2012年6月

人机交互发展趋势

交互变革：人学习机器 => 机器学习人，优质内容服务，要通过自然人机交互到达更多普通用户。
人机交互方式面临变革机会，影响物联网的规模增长，也将带来产品、服务和商业模式的变革。



人工智能 (AI) : 机器实现人脑功能

- 知识 / 规则 / 模板
- 符号推理

将人类知识赋予机器去解决问题。

机器学习 (ML) : 信息论与概率统计

- 有监督 : Supervised
 - 分类 : kNN / SVM / MLP...
 - 回归 : LR / SVR...
- 无监督 : Unsupervised
 - 聚类 : K-means...
 - 降维 : PCA / LDA...

让机器通过数据学习变得越来越强大。

深度学习 (DL) : 基于大数据的表征学习

- DNN / CNN / RNN / LSTM / Attention ...
- DRL / GAN / ...

语音 / 图像等领域突破性进展, 进入产业化阶段。

1950's

1980's

2010's

感知 (Perception)

- 听觉：信号 / 语音 / 声纹 / 情绪等
- 视觉：图像 / 人脸 / 视频等

表达 (Expression)

- 语音合成 / 翻译 / 表情等
- AR / VR 等

认知计算 (Cognitive Computing)

- 语言：意图 / 信息 / 情感等（感知理解和生成结果表示）
- 知识：实体 / 关系 / 结构 / 逻辑，用户画像 / 语境信息等
- 思维：理解 / 推理 / 决策 / 规划 / 生成 / 自学习等

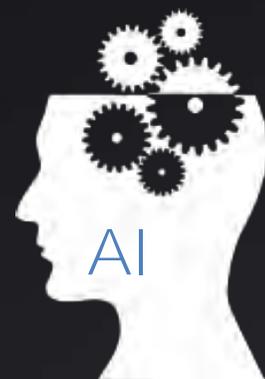
基础架构：机器学习 + 超级计算 + 大数据

人工智能 (AI) 发展现状：

- 计算：碾压人类 (AlphaGo)
- 感知：接近人类 (适应性)
- 表达：有待完善 (丰富自然)
- 认知：相去甚远 (知识 / 思维)
- 意识：尚不存在 (自我意图)

人工智能 (感知 / 认知 / 表达)，是以**机器学习、超级计算和大数据**为坚实基础的。

深度学习目前在感知和表达技术方面都取得突破，在认知计算方面还有待更大突破。



1、产业全面升级势在必行

- 用户需求：智慧生活
- 企业需求：智能制造 + 智慧服务
- 政府需求：智慧城市

2、技术基础持续高速发展

- AI 技术突破：深度学习
- 大数据积累 (TB / PB)
- 超级计算普及 (T / P FLOPS)

产业需求驱动，技术日趋成熟，企业成为主导力量，商业、管理、思维等模式有待变革。



智能语音交互技术发展

云知声：专注物联网人工智能服务

51CTO WOTI

以 AI 基础架构和技术为支撑，以语音交互为切入点，推动 AI 在物联网的产业应用，构建大数据闭环，为生产制造、内容服务和公共基础设施优化提供科学依据和解决方案。



云知声技术创新和产业实践历程

云知声云端布局：

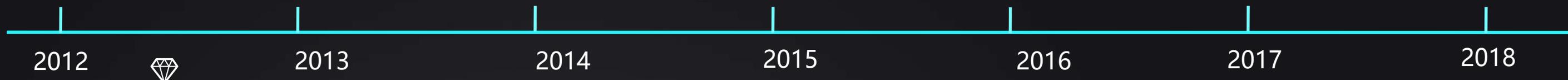
- 开放语音云平台发布
- 率先升级深度学习 ASR 引擎
- GPU 集群搭建
- 语义团队组建

物联网“云端芯”战略布局：

- 远讲 / 降噪 / 唤醒技术
- 智能家居 IVM 方案，支持5m远讲
- 智能车载 AIUI 方案
- \$50M B轮融资

认知和计算架构升级：

- Sextant 超算平台和语用计算发布
- 终端渗透：DNN @ Cortex - M
- 智能家居规模化量产，医疗应用落地
- 上榜科技部“独角兽”企业，B+ 融资



云知声成立

- 顶尖研究所 / 高校
- 10 年技术积累
- 5 年工业界经验
- AI 产业升级挑战

终端交互数据闭环：

- LeTV 智能交互方案
- DNN @ Cortex - A 移植
- 语义云平台发布
- 亿级 A 轮融资

深度学习全面升级：

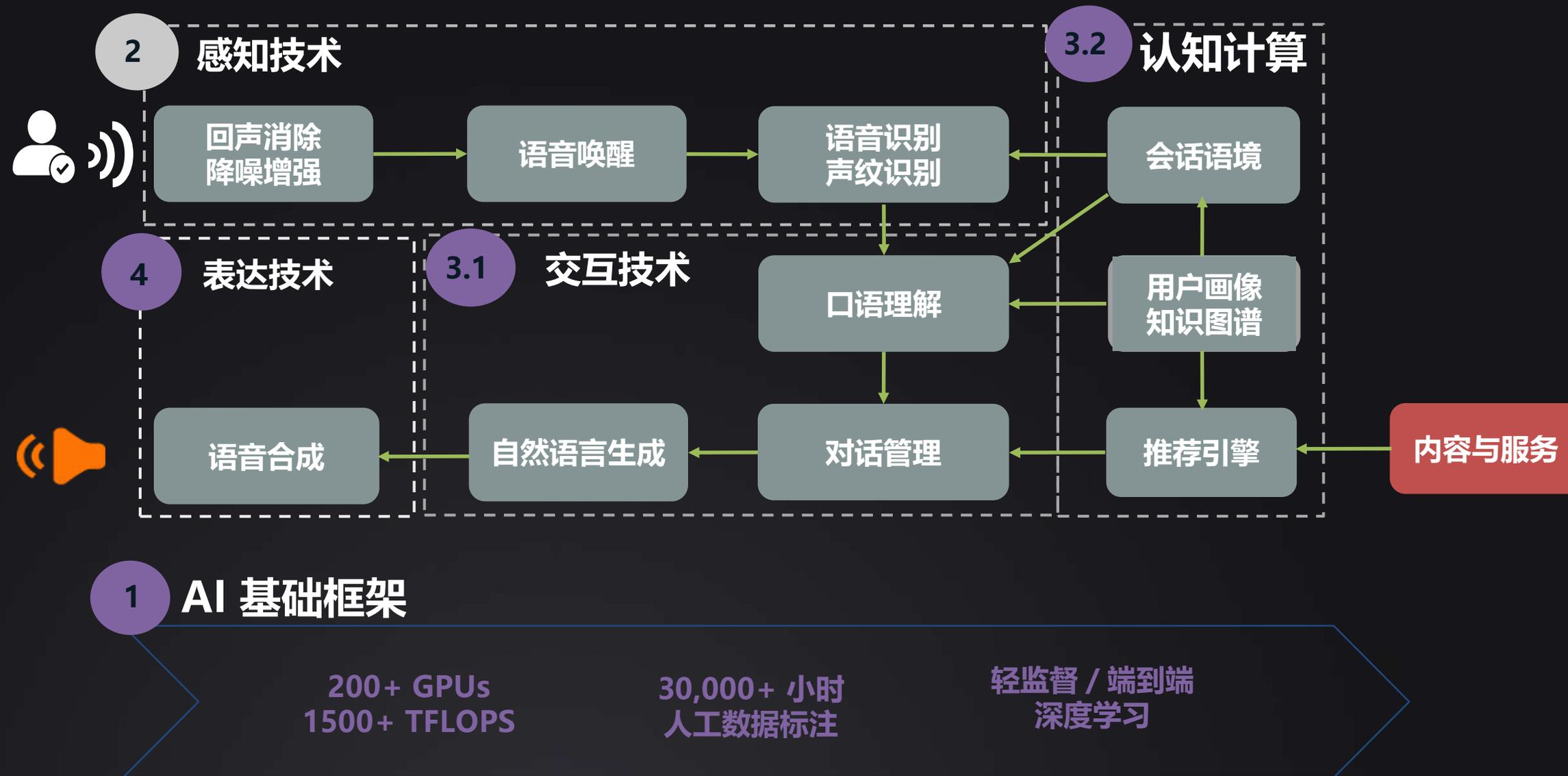
- Seq2Seq NLU，LSTM ASR/TTS
- DNN @ AIChip 发布
- 车载后装市场占有率第一
- 福布斯成长最快科技企业

技术拓展 - 产品闭环优化：

- AIOS & Pandora 发布
- 图像 / 翻译技术扩展
- 联合实验室深度合作
- AIChip 升级 (UniOne)

面向物联网，完善 AI 技术体系，以“云端芯”战略和产品方案，在家居、车载、医疗、教育等领域实现产业化。

云知声智能语音交互技术体系：感知 + 认知 + 基础架构



关键技术突破：远讲和降噪

双麦 (2Mic) 降噪：

- 180度声源定位、盲源分离、语音增强
- 回声消除、混响和噪声抑制
- 模型深度适配

优势：

- 结构灵活性
- 硬件一致性要求低
- 功耗和成本低

劣势：

- 降噪技术难度大，
- 需要更强建模能力

国际代表：

Google Home 技术方案



多麦 (4Mic+) 阵列：

- 360度声源定位、波束形成、语音增强
- 回声消除、混响和噪声抑制
- 后端模型适配

优势：

- 信号降噪技术相对成熟

劣势：

- 结构不灵活
- 硬件一致性要求高
- 资源、功耗和成本高

国际代表：

Amazon Echo 技术方案



关键技术突破：低功耗唤醒和识别

DNN @ X86 : 2012

- 2GHz+ / 1GB+
- Windows / Linux

DNN @ Cortex-A : 2013

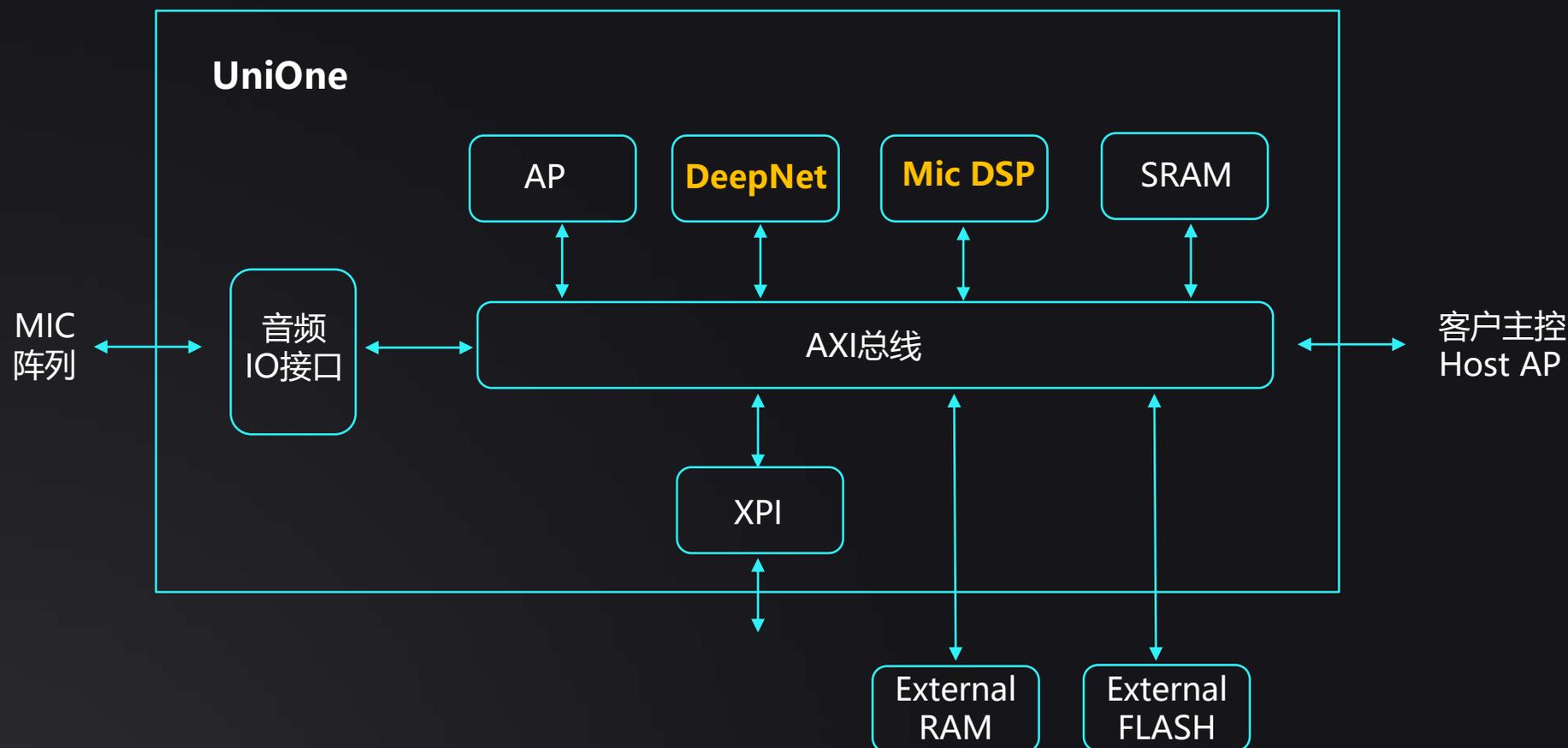
- 1GHz+ / 1MB+
- Linux / Android

DNN @ ASIC : 2015

- ARM9 + DNN DSP
- Linux

DNN @ Cortex-M : 2016

- ~100MHz / 100KB
- RTOS



针对垂直领域的识别优化

51CTO WOTI



医学病历语音输入，准确识别大量专业词汇、英文、符号等



智能家居的自然语言交互（语音 + 语义）



数千万 POI 的识别优化，离线在线识别结合



口语评测技术全面升级

高质量、多音色、个性化语音合成技术

51CTO WOTI



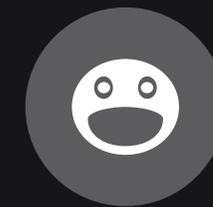
高质量

- 基于CRF/LSTM 前端信息预测
- 基于HMM/LSTM后端声学参数预测
- 基于 DNN 声学参数增强
- 高质量合成器



多音色

- 播音风格男声
- 播音风格女声
- 甜美风格女声
- 台湾腔女声
- 可爱童声
- 持续增加中.....



个性化

- 使用较少数据 (~15 分钟)
- 无监督快速构建合成系统



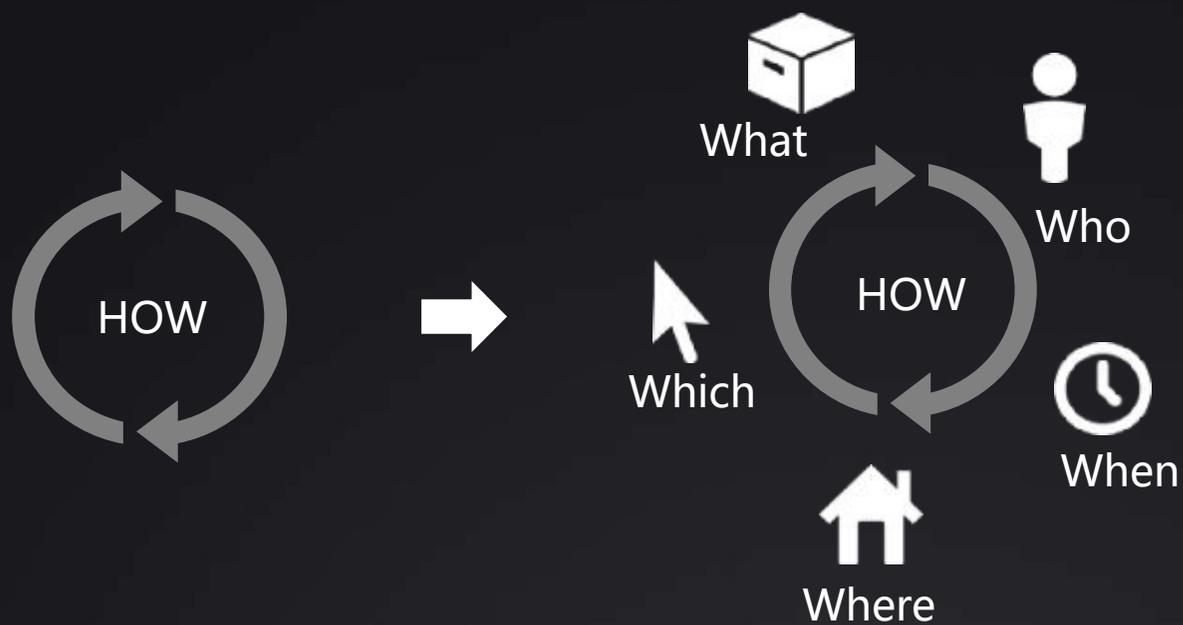
会话式交互技术：从理解到生成



语用计算：融入语境信息的端到端(Seq2Seq)深度学习对话系统

51CTO WOTI

语用计算是目前最先进的自然语言处理技术，系统结合多种语境信息，包括物理语境，言语语境和知识语境，支持基于语境的口语理解，生成和人机交互框架。2016年8月份，云知声国内首家上线融入语境信息的端到端(Seq2Seq)对话系统，解码速度比谷歌开源系统 TensorFlow 快 **15** 倍。



语义理解

语用计算

输出字面意义信息或动作

基于用户结合环境和设备输出有用的信息或动作

聊天		你好		
			下午好	
问答		明天青龙峡天气怎样？	天气很好，晴，20-28度，微风。	
聊天		怎么这么冷？	山区温度低，记得多带些衣服~	
任务		订一家那附近的农家院。	要什么价位的？	
任务		300~500元一晚的。		

基于语用计算的多轮对话交互



智能交互的物联网应用实践

互联网发展的未来：自然交互 + 精准 / 个性化服务

51CTO WOTI



智能制造



智慧城市



智慧生活

物联网智能交互的挑战

智能交互：以语音交互为主的多模态自然人机交互，让用户能够自然和便捷的通过物联网设备获取内容服务。

智能服务：基于大数据的精准和个性化的内容服务，并反向优化生产制造、内容服务和公共基础设施等。



云知声的“云端芯”战略和产品体系

51CTO WOTI

AI 核心技术
学习、感知、认知

AIOS (端)
终端交互体验
应用领域：车载 / 家居



AI Service (云)
云端认知和智能服务
应用领域：IoT / 医疗 / 教育

AI Chip (芯)
终端感知和本地智能
应用领域：家居 / 车载

云知声智能家居语音交互方案：IVM (Intelligent Voice Module)

51CTO WOTI



长虹智能空调



美的空调



格力空调



海尔空调



华帝抽油烟机



乐视超级电视



康力优蓝机器人

数十家家电厂商合作

十余款标杆产品落地

百万台出货量

数千万覆盖人群



云知声智能家居中控方案：AIOS@Home (Pandora)

Pandora 解决方案可以应用在各个领域不同形态的AI产品中，构建以智能语音交互为入口的万物智联生态



云知声车联网语音交互方案：AIOS@Car

51CTO WOTI

拾音降噪、语音交互、云端计算和内容服务一揽子解决方案。



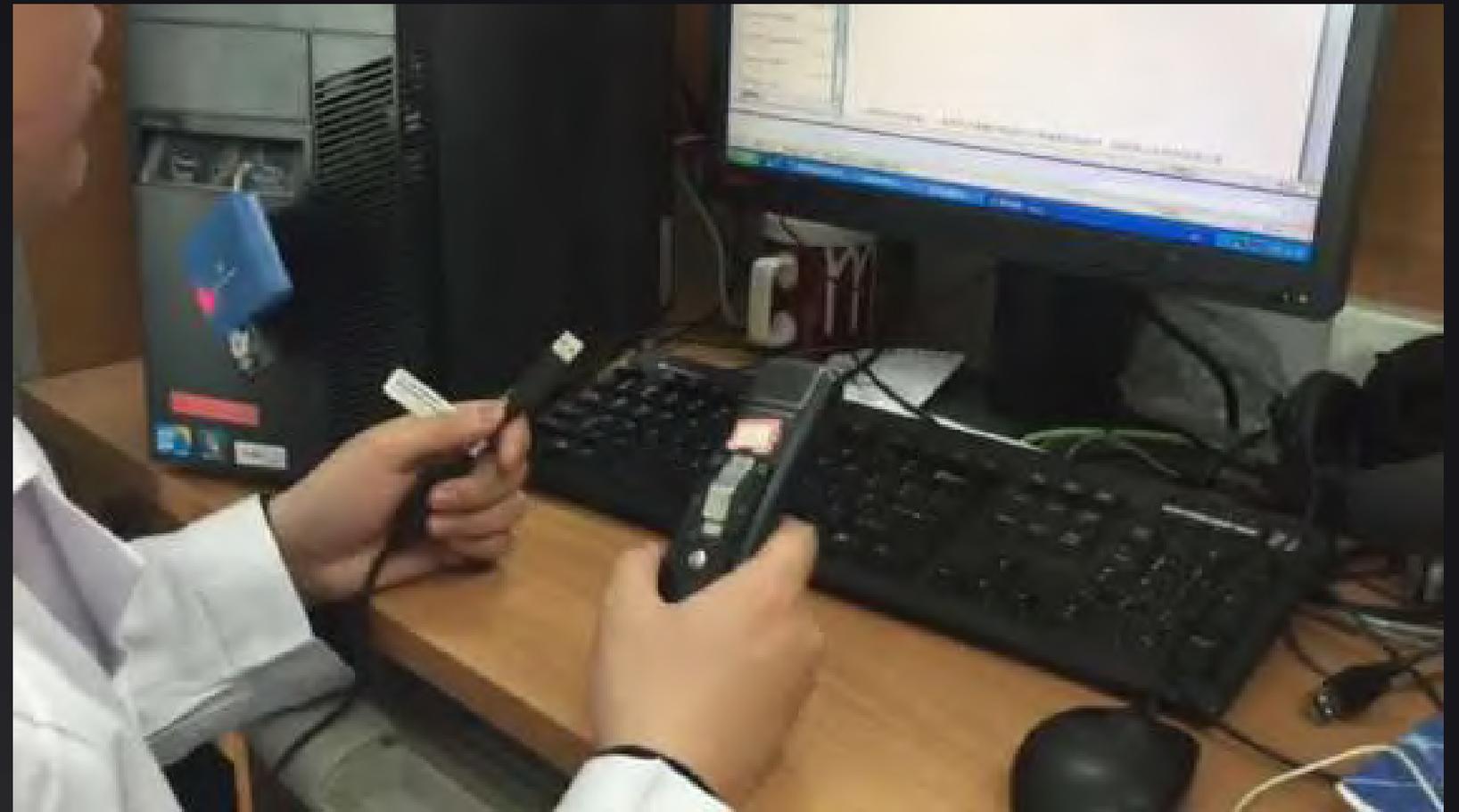
场景化识别

软硬降噪方案

软件AEC方案

云知声智能云服务方案：AI Service@HealthCare / Education

51CTO WOTI



云知声物联网生态体系

51CTO WOTI

以 AI 核心技术为基础，以“云端芯”产品体系为支撑，面向智慧生活和智慧服务，为智能终端产品提供智能服务解决方案。



Athena

智慧大脑



产品体系



应用于两大垂直方向



为智能终端产品提供智能服务解决方案

- 物联网与人工智能的结合是大势所趋，智能语音技术迎来产业化机遇和挑战
- 语音技术取得突破性进展，但产业化仍面临噪声、口音、领域、功耗、成本等挑战
- 未来更大的挑战在认知计算（语言、知识、思维能力），是交互和服务升级的关键
- 开放的“技术 - 产品 - 服务”生态体系，有利于 AI 产业的快速发展



THANK YOU