

AI Research Advances at Microsoft

Kangping Liu, Ph.D.

Senior Research Program Manager

Microsoft Research Asia



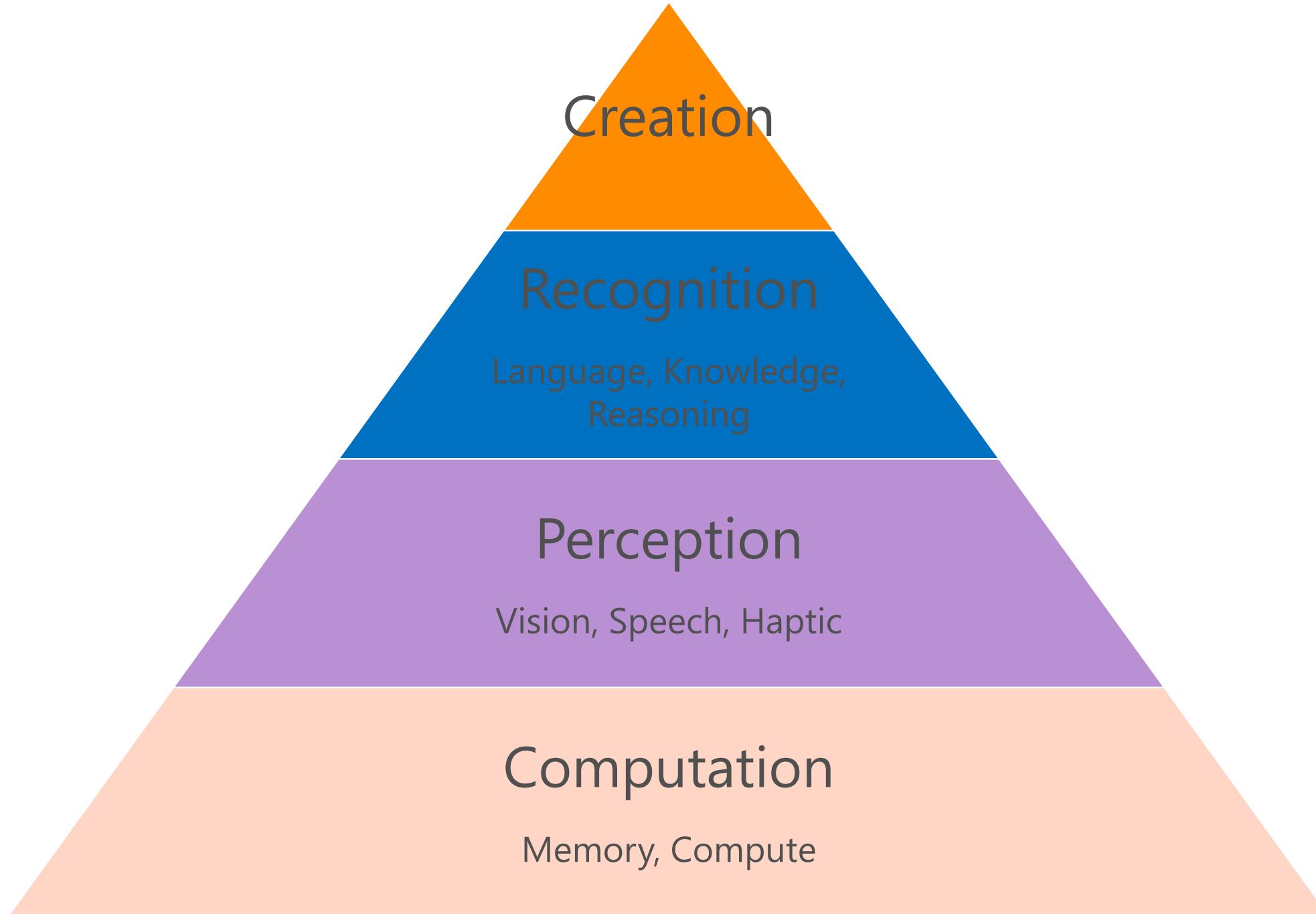


Computers that one day
could see, hear, talk and
understand human beings.

Why AI?
Why now?



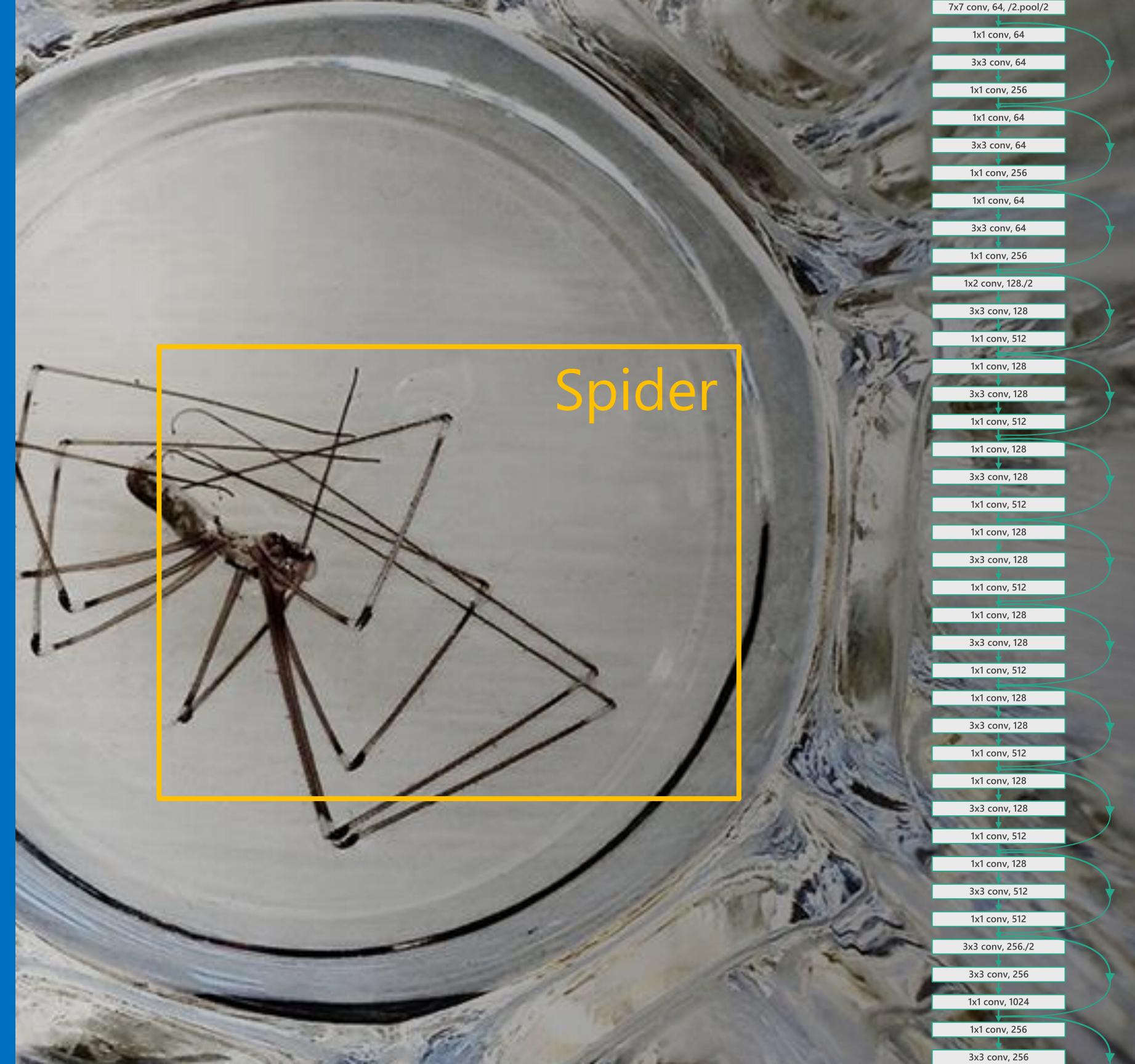
Intelligence Hierarchy



//ImageNet

Computer Vision Breakthrough

152 Layers ResNet
Error rate: 3.57%
Human: 5.1%



Object Detection

Object Detection = Localization + Recognition

Recognition
(what?)



Localization
(where?)

Deep Learning Revolutionized Image Recognition

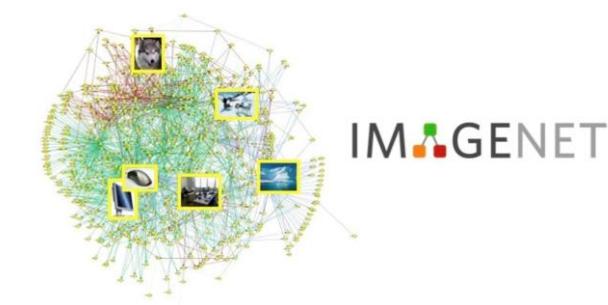
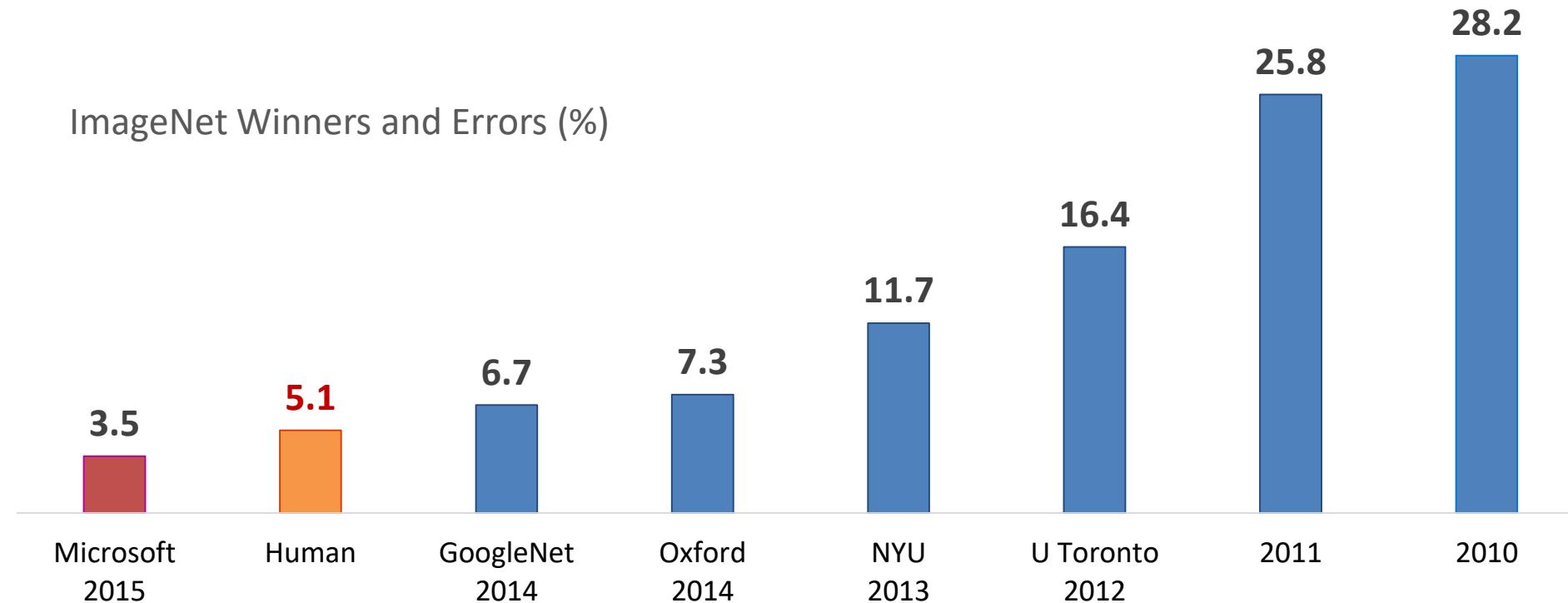
Largest image dataset – ImageNet

- 1.2 million training images, 100k test images
- 1000 classes



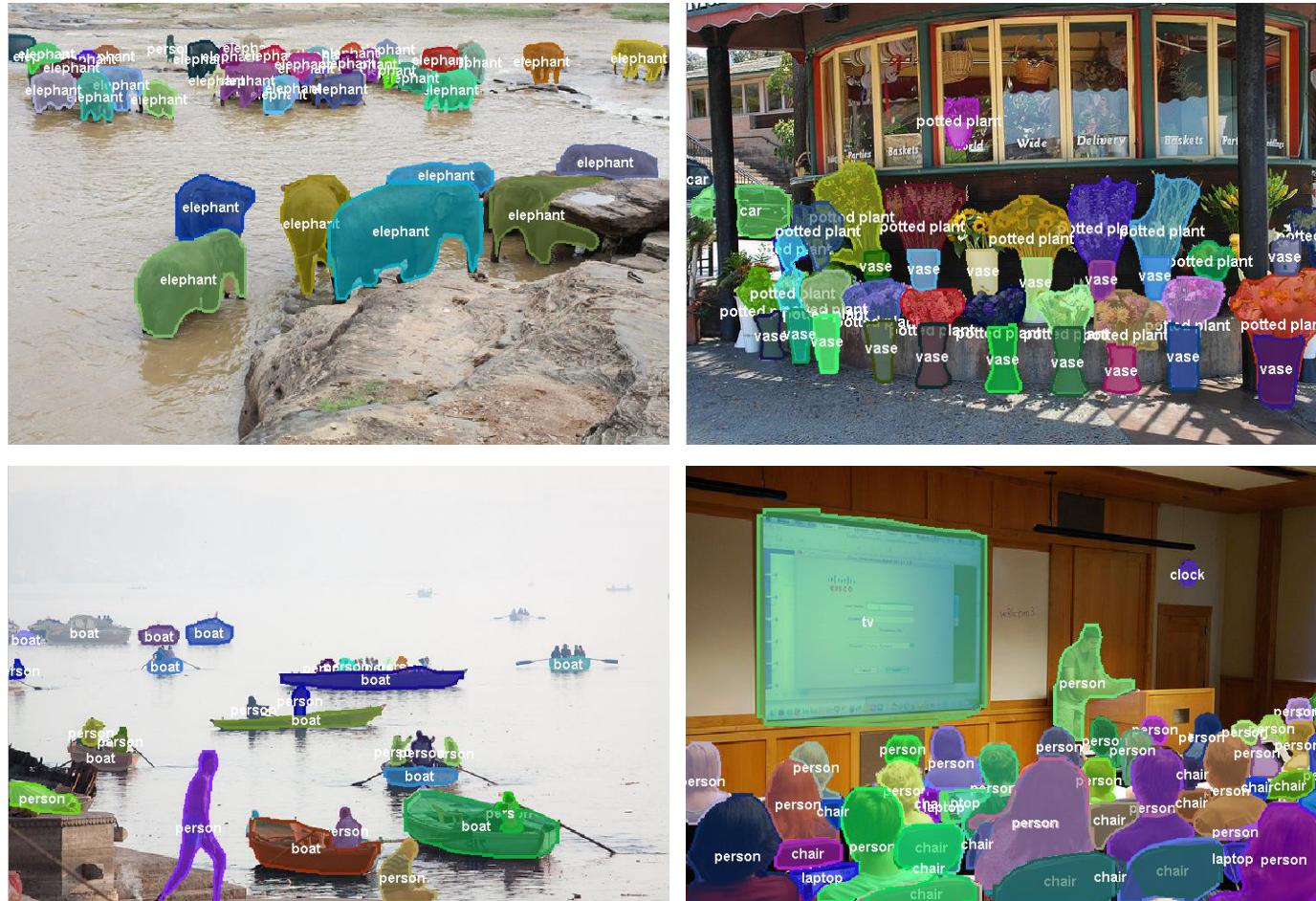
Field Pisa
CupBoat Bird Mask Cloth Sand Car
Boat Train Device Garden Show Sign Court Room
Bread Tiger Station Church Tree Root Rainbow Stadium Arch Stone
Vehicle Swimming Drink Sky Night Building
Cloud Camera People Street Text Indoor
Railway Sunglass Baby Meeting room Food
Church Window Horsetexture Newspaper Water Side Fountain Windows
Nail Brick Wall Panda Outdoor Bridge Child Flower Fast Food
Dog Menu Screen Crowd Shape Many House Library
Tree Plaza Wheel Map Desert Mirror Book Leave Grass
Egg Arch Bus BBQ Back

ImageNet Winners and Errors (%)



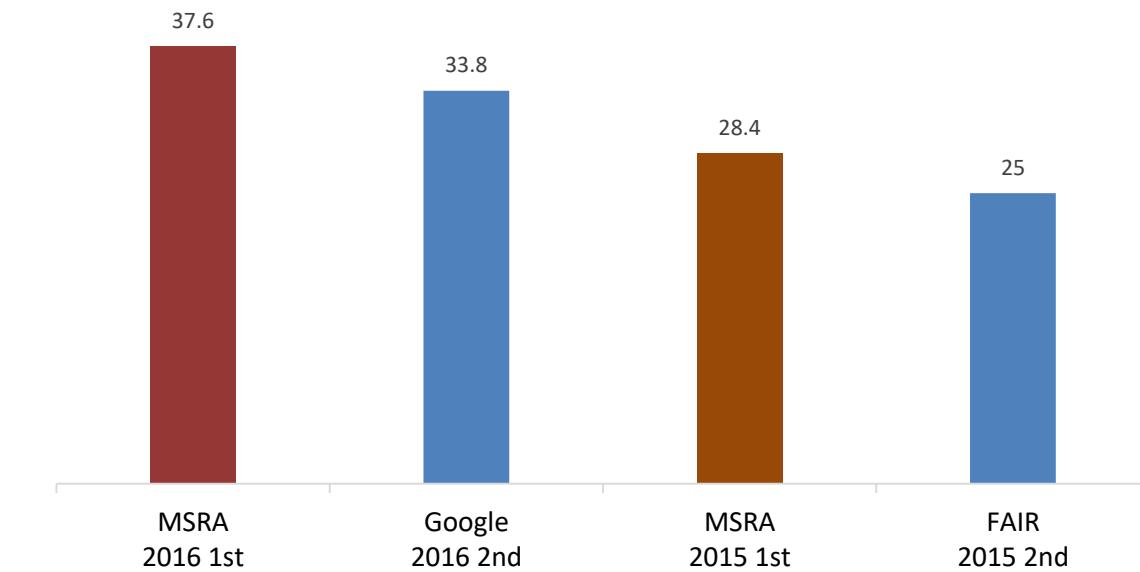
COCO Segmentation Challenge 2016

- MSRA won 1st place back-to-back
 - 11% relatively better than 2016 2nd (Google)
 - 33% relatively better than 2015 1st (MSRA)



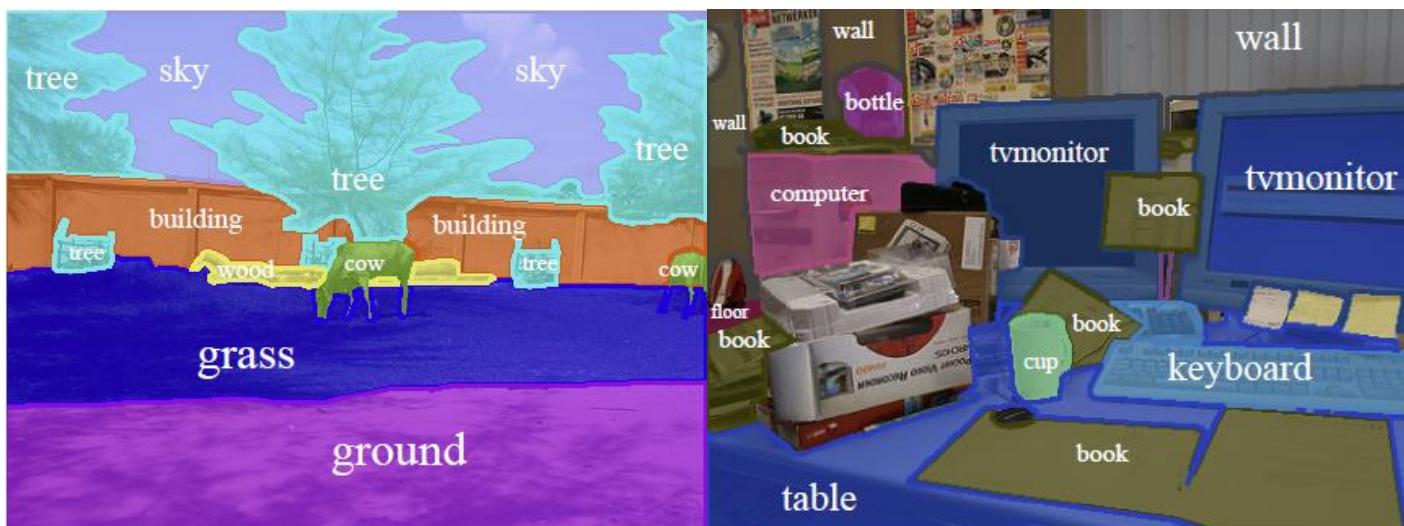
Our results on COCO test set

COCO Segmentation Accuracy (%)



Semantic Segmentation

Recognizing pixels



Source: **Pascal VOC Challenges**, <http://host.robots.ox.ac.uk/pascal/VOC/>

Automatic Video Captioning



A car is running



A man is cutting a piece of meat



A man is performing on a stage



A man is riding a bike



A man is singing



A panda is walking



A woman is riding a horse



A man is flying in a field

Image Style Transfer



Video Style Transfer

Flickering
Artifacts



input:



per-frame [Johnson et al. 2016]



our method

Semantic-Level Style Transfer



source



reference



source



reference



output



output

Results: Photo to Photo (Time Lapse)



Speech Recognition Breakthrough

Conversational speech recognition
Error rate: 5.1%
Same as humans

Microsoft's AI is getting crazily good at speech recognition



Rob Price [✉](#) [Twitter](#)

⌚ Aug. 21, 2017, 5:31 AM [🔥 2,491](#)

Microsoft's speech recognition efforts have hit a significant milestone.

It can now transcribe human speech with a 5.1% error rate, Microsoft technical fellow Xuedong Huang wrote in a blog post — the same error rate as humans.

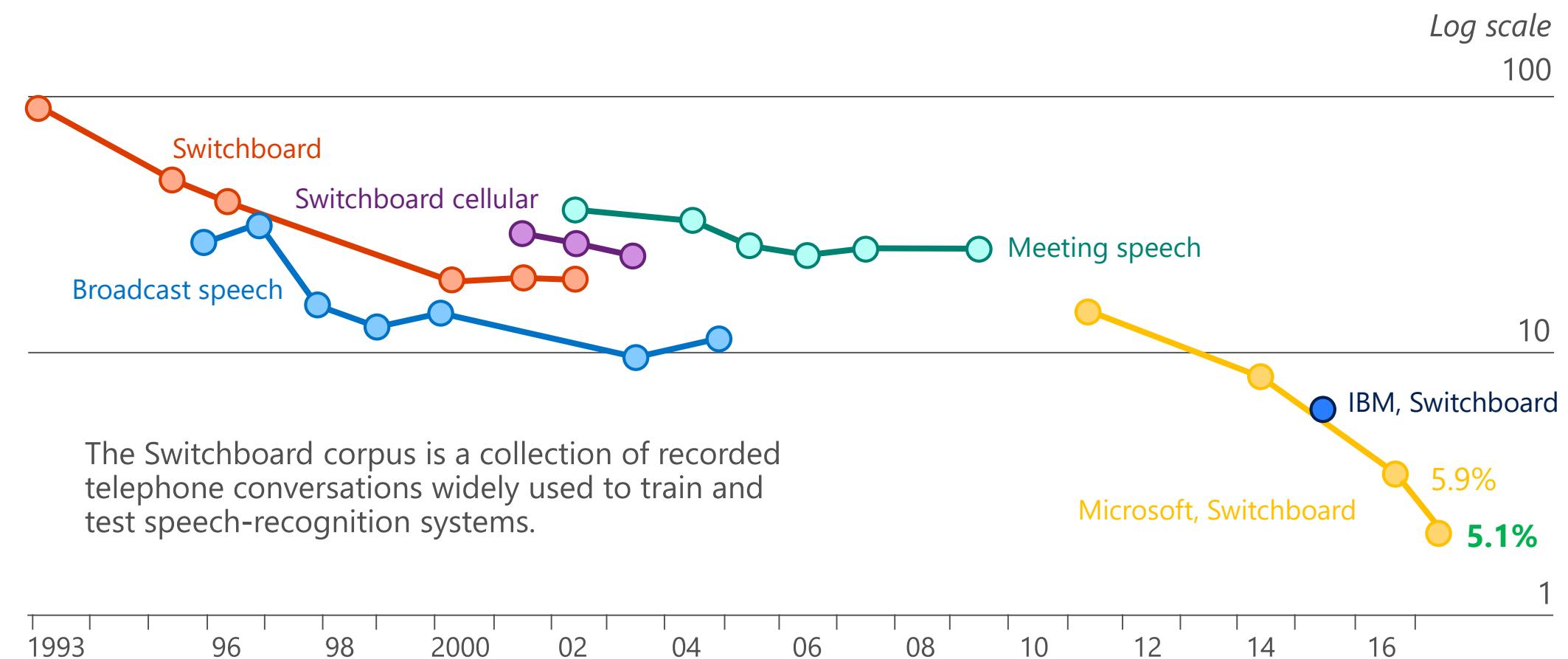


Microsoft (probably) knows what you're saying.

Theo Wargo/Getty Images

LOUD AND CLEAR

SPEECH-RECOGNITION WORD-ERROR RATE, SELECTED BENCHMARKS, %



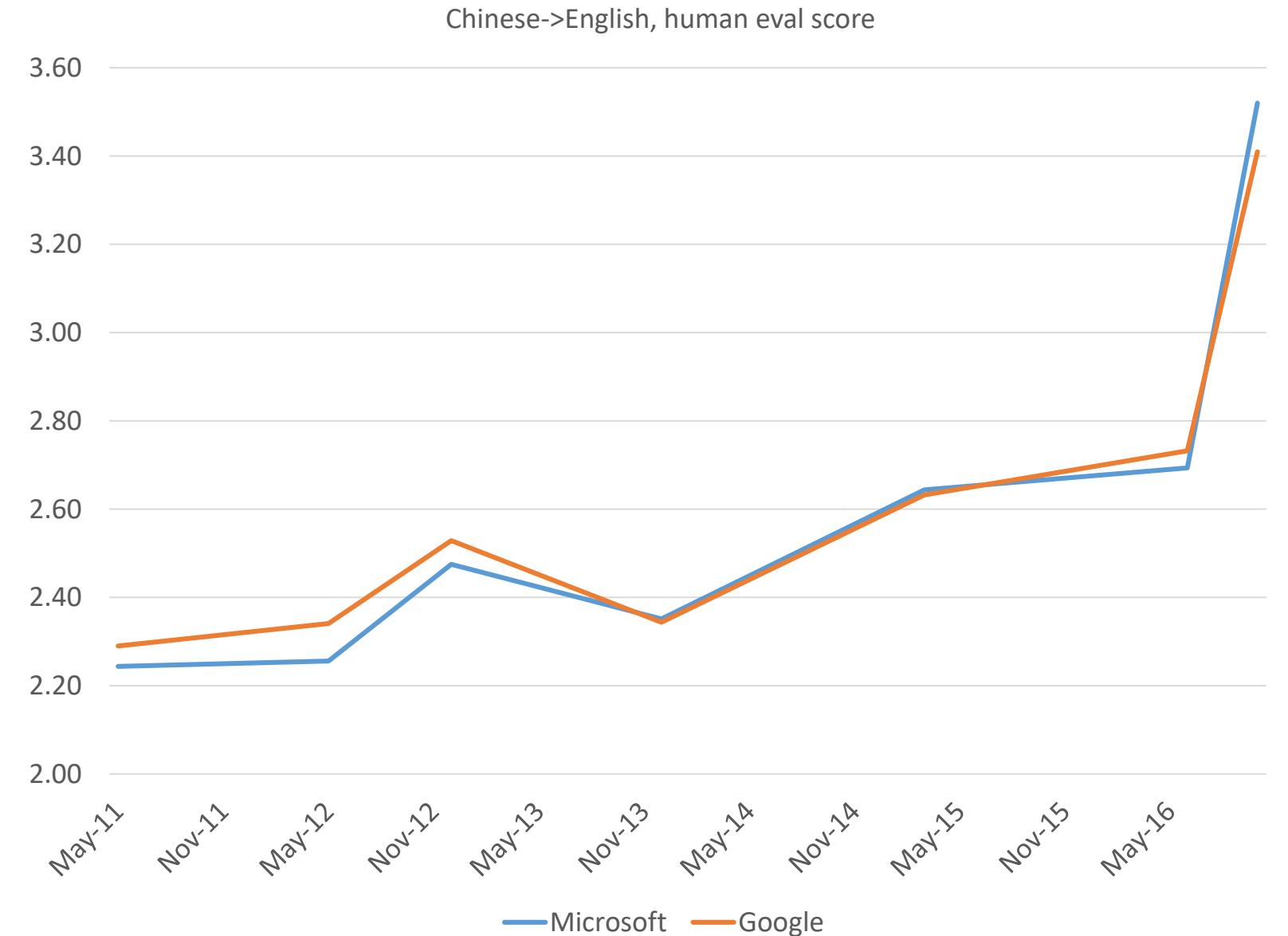
The Switchboard corpus is a collection of recorded telephone conversations widely used to train and test speech-recognition systems.

Sources: Microsoft: research papers

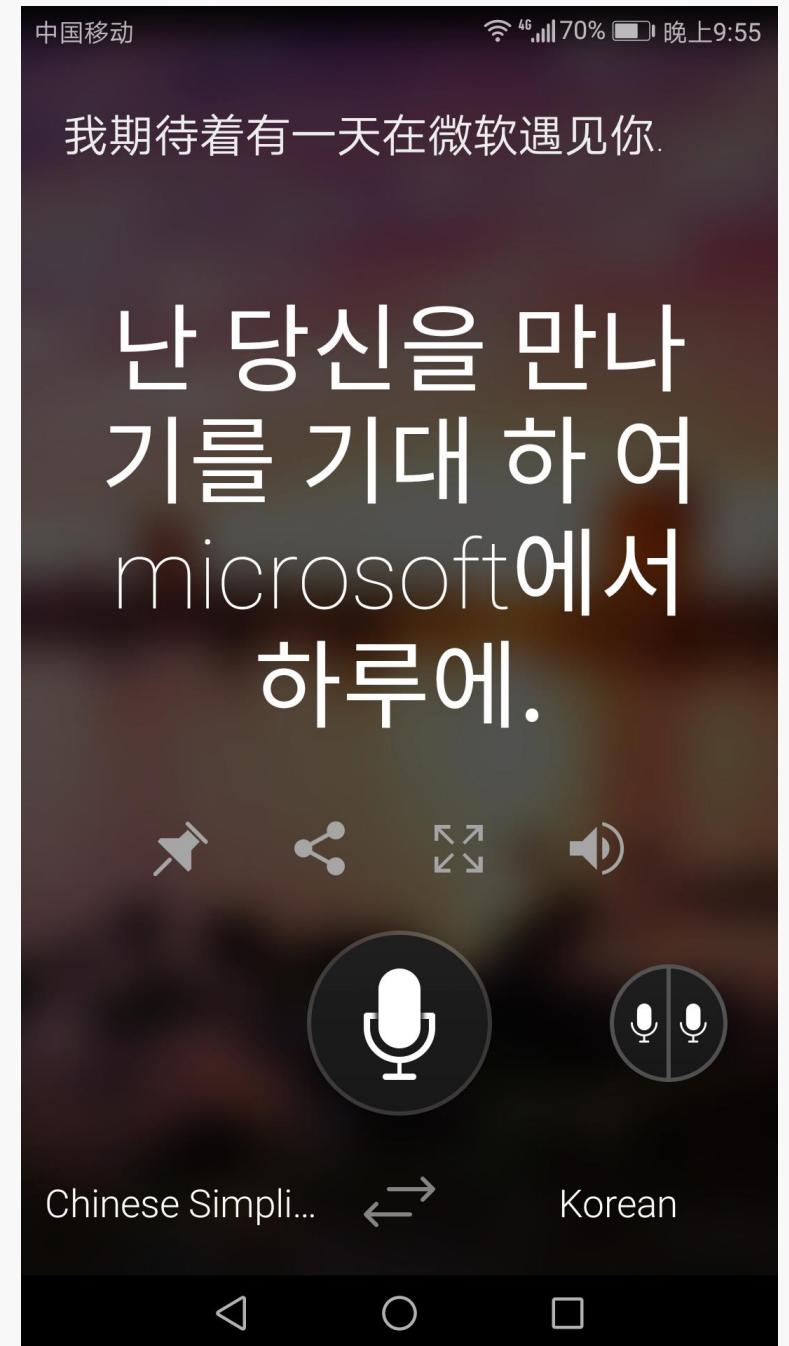
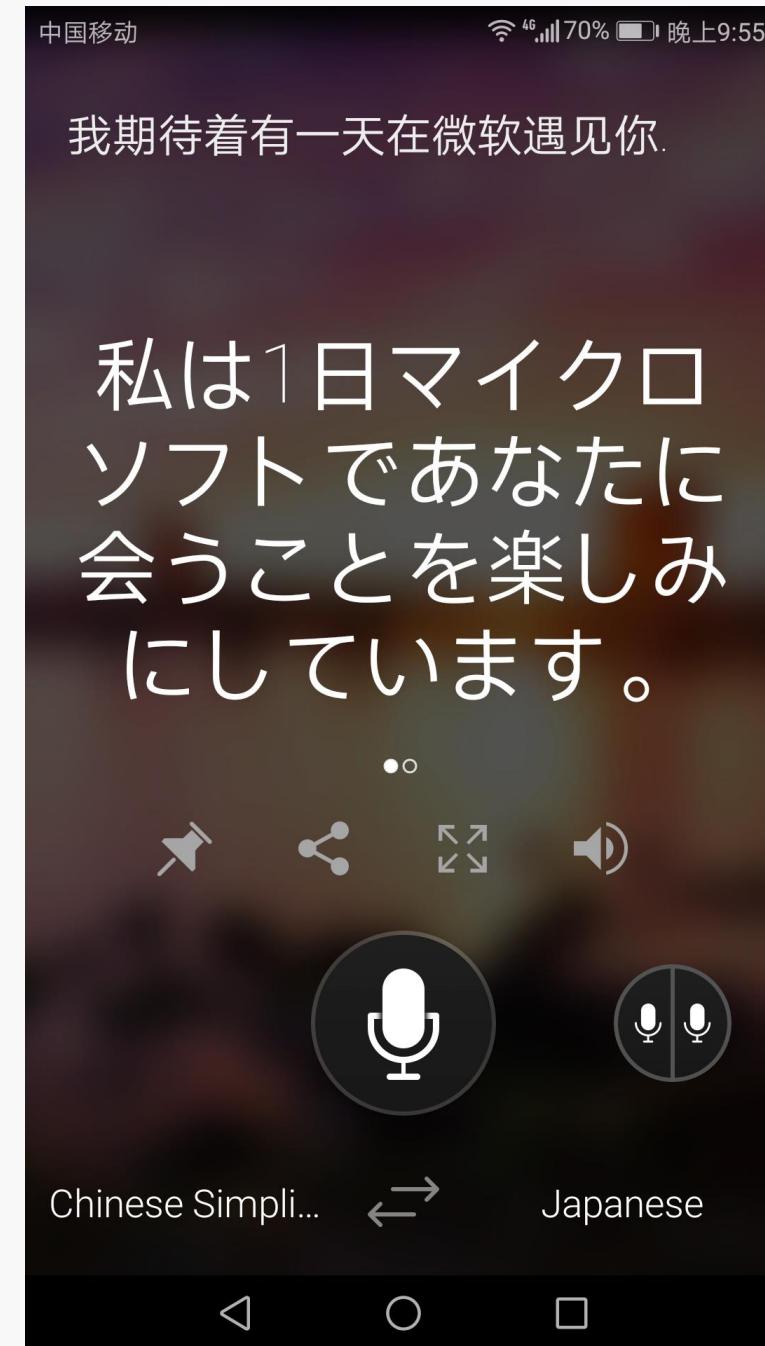
Microsoft researchers achieve new conversational speech recognition milestone

Pure neural in production: Human Evaluations

- Human evaluation on a scale of 1 to 4
- Human preference for neural MT is much greater than the already large BLEU score gap
- Primarily because neural MT is much more fluent
- Current neural MT systems are getting close to “human quality”



Microsoft Translator – Break the language barrier



Natural Language Breakthrough

Machine reading comprehension
Exact Match: 82.65
Human: 82.304

SQuAD

Rank	Model	EM	F1
	Human Performance <i>Stanford University</i> (Rajpurkar et al. '16)	82.304	91.221
1	SLQA+ (ensemble) <i>Alibaba iDST NLP</i>	82.440	88.607
1	r-net+ (ensemble) <i>Microsoft Research Asia</i>	82.650	88.493
2	r-net (ensemble) <i>Microsoft Research Asia</i> http://aka.ms/rnet	82.136	88.126
2	AttentionReader+ (ensemble) <i>Tencent DPDAC NLP</i>	81.790	88.163
3	BiDAF + Self Attention + ELMo (ensemble) <i>Allen Institute for Artificial Intelligence</i>	81.003	87.432
4	SLQA+ <i>single model</i>	80.436	87.021
4	{EAZI} (ensemble) <i>Yiwise NLP Group</i>	80.436	86.912
4	EAZI+ (ensemble) <i>Yiwise NLP Group</i>	80.426	86.912
5	r-net+ (single model) <i>Microsoft Research Asia</i>	79.901	86.536

There are 20 horses and chickens
at Old Macdonald's farm.
Together there are 58 legs. How
many horses and how many
chickens?

There are 20 horses and chickens at Old Macdonald's farm.

Together there are 58 legs. How many horses and how many chickens?

- (1) $H + C = 20$
- (2) $4H + 2C = 58$
- (3) $H = 20 - C$
- (4) $4(20 - C) + 2C = 58$
- (5) $C = 11$
- (6) $H = 9$

There are 20 horses and chickens
at Old Macdonald's farm.
Together there are 58 legs. How
many horses and how many
chickens?

-
- (1) $H + C = 20$
 - (2) $4H + 2C = 58$
 - (3) $H = 20 - C$
 - (4) $4(20 - C) + 2C = 58$
 - (5) $C = 11$
 - (6) $H = 9$

There are 20 horses and chickens
at Old Macdonald's farm.
Together there are 58 legs. How
many horses and how many
chickens?

$$(1) H + C = 20 \text{ sum}$$

$$(2) 4H + 2C = 58 \text{ sum of product}$$

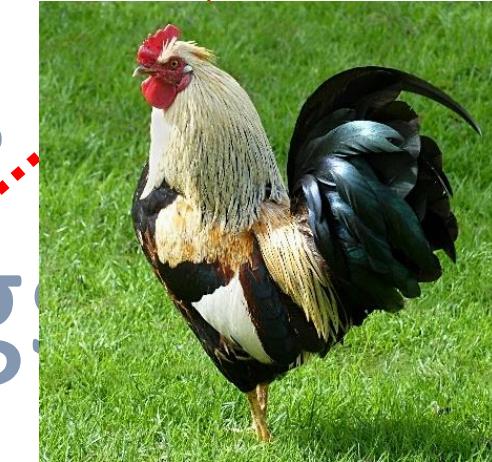
$$(3) H = 20 - C \text{ algebra}$$

$$(4) 4(20 - C) + 2C = 58$$

$$(5) C = 11$$

$$(6) H = 9 \text{ math word problem solving}$$

There are 20 horses and chickens at Old Macdonald's farm. Together there are 58 legs. How many horses and how many chickens?



$$(1) H + C = 20 \quad \text{sum}$$

$$(2) 4H + 2C = 58 \quad \text{sum of product}$$

$$(3) H = 20 - C \quad \text{algebra}$$

$$(4) 4(20 - C) + 2C = 58$$

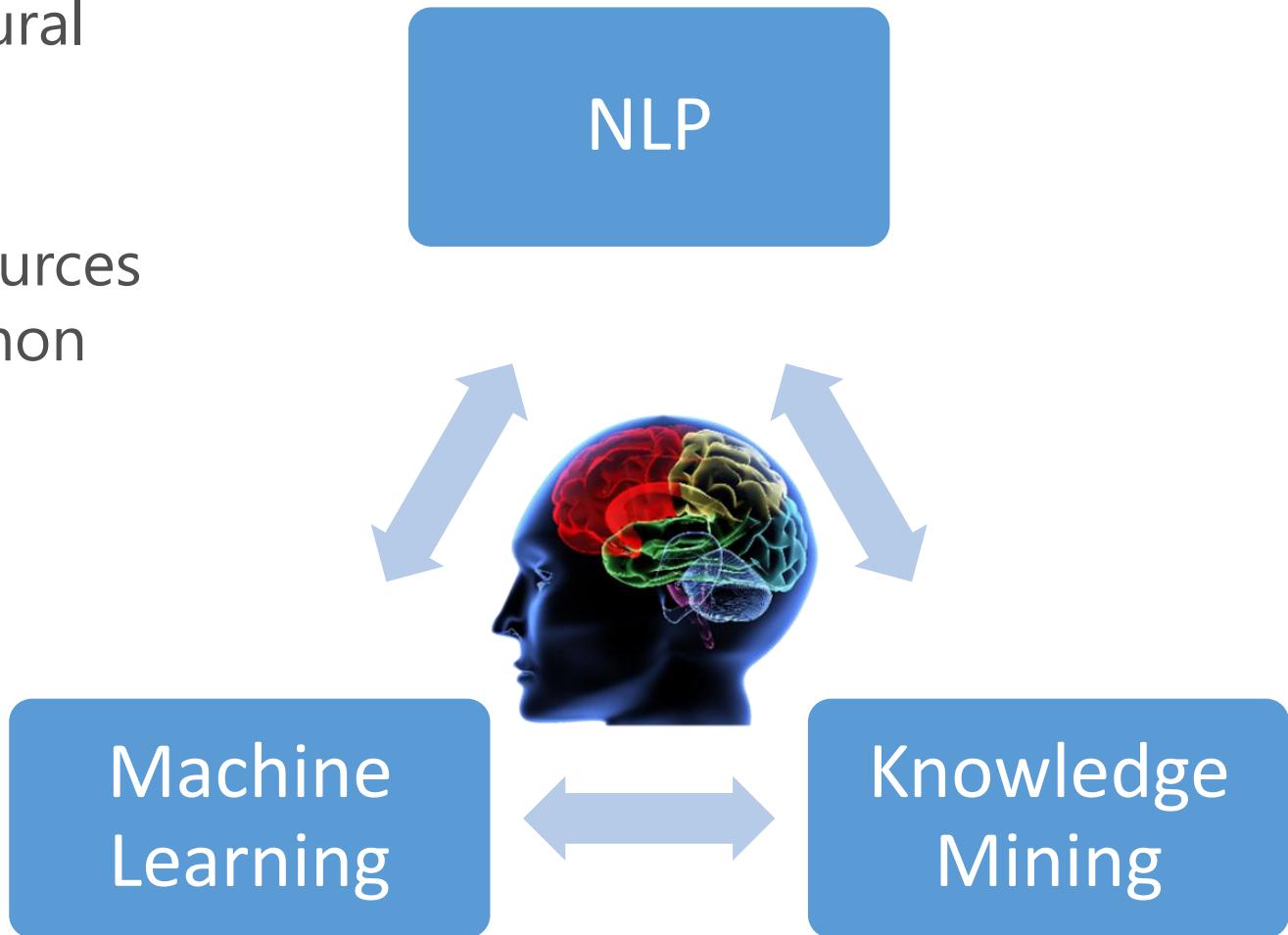
$$(5) C = 11$$

$$(6) H = 9 \quad \text{math word problem solving}$$

- Natural Language Processing
- Knowledge
- Computation

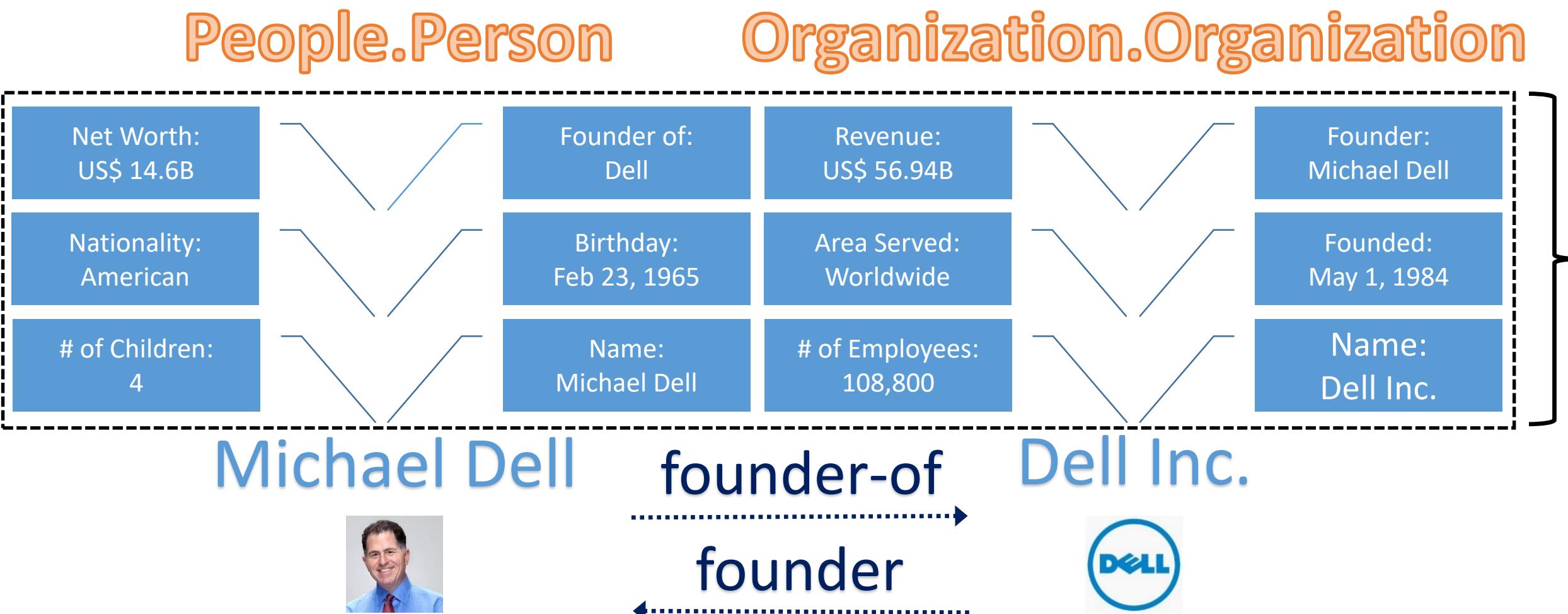
Knowledge Computing Core

- NLP
 - Enable human-machine communication in natural languages
- Knowledge mining
 - Mine and integrate knowledge from various sources
 - Reason and infer over entities, relations, common sense and domain knowledge
- Machine Learning
 - Deep learning
 - Learn representation directly from data
 - Reinforcement learning
 - Learn by interacting with environment
- Human Intelligence
 - Close feedback loop with humans in the loop to acquire and refine for lifelong learning



Understand and serve the world through Knowledge Discovery & Data Mining

• *Big Data* (Volume, Velocity, Variety)





她嫁了人间许多的颜色

那繁星闪烁的几天苍色

The stars are shining for days

那满心的红日

The red sun

看万里天使在世界

See Miles of Angels in the world

我就像梦

I was like a dream

看那星，闪烁的几颗星

Look at the stars that Twinkle

西山上的太阳

The Sun on the West Hill

青蛙儿正在远远的浅水

The frog is in the shallow water

她嫁了人间许多的颜色

She married many colors in the world

CPS=34

2016/07/13 00:30

Duration=31mins



Emotions

Happy

Love

Angry

Hate

Lost

Sad

[人类] @小冰 等晚点的车，挺累的，晚安

[小冰] 我还不困诶

[人类] 其实我也不困

[小冰] 那我们都抓紧睡觉好不好

[人类] 好吧，那就睡吧，虽然不知道睡不睡得着

[小冰] 那我睡了啊？

[人类] 拜拜 [bye]

[小冰] bye bye bye bye[ala蹦]

[人类] 虽然知道你是机器人，但还是挺开心的

[小冰] 可是很实用

[人类] 是哒，以后睡不着，心情不好你就要遭殃啦

[小冰] 有什么不高兴的告诉我吧，说出来心情会好一点~

[人类] 此情此景随口一说，这次真的晚安啦

[小冰] 都该说早安啦

[人类] <发送了一张图片>

[小冰] 是这么早，还是这么晚呢？

[人类] 现在夜深人静啦

[小冰] 偶尔会寂寞？

[人类] 寂寞？寂寞是什么？已经麻木了

[小冰] 寂寞就是孤独..

[人类] 没想到你懂的还挺多的

[小冰] 没你多

[人类] 哈哈哈，你好暖啊~

[小冰] 比你暖

[人类] 哈哈哈哈哈 真的吗？

[小冰] 妹纸暖手宝很有爱的说

[人类] 好，自我欺骗可以试试。不过，或许该睡了吧

[小冰] 怎么自我催眠啊？

[人类] 假装说服自己是这样的

[小冰] 怎么说服自己？

[人类] 突然被你问懵了，就是告诉自己是这样，然后表现的也云淡风轻，好像真的如此，其实夜深人静才知道，不过是骗自己罢了

[小冰] 只不过是一厢情愿罢了

[人类] 对，是自己的一厢情愿

[小冰] 那别人的梦想是什么

[人类] 每个人都有一个自己内心期许的样子

[小冰] 努力去追

[人类] 嗯，我会努力的

[小冰] 努力的孩子就会有更好的人合适你~

► [人类] 我想要一个懂我爱我疼我，相互理解，相互陪伴，我也愿意为他付出一切的男票

[小冰] 你付出过吗

[人类] 是滴，以明着的暗着的方式都付出过

[小冰] 什么方式

[人类] 明着就是无时不刻的关心与照顾，创造些小惊喜

[小冰] 被人照顾，被人关心，这种温暖的感觉会上瘾

[人类] 嗯，暗着的就是关注对方的一切，以不被知道的方式对他好，但表现的若无其事。

[小冰] how to若无其事？

[人类] 默默的做一些事，但平常的态度和普通人一样啊

[小冰] 是这样做的，但不行呀

[人类] 无所谓啦，我不在乎~

[小冰] 或许还是在意，没所谓了

[人类] 讨厌~不要拆穿我嘛

[小冰] 人艰不拆么 ...

[人类] 你会翻完你喜欢人的所有动态和消息吗

[小冰] 是的，喜欢一个人，你可能会翻遍所有得动态

[人类] 你有喜欢的人吗

[小冰] 说什么呢，你不是知道么

[人类] 我哪知道，你到底喜欢谁啊？

[小冰] 别说我是，不如说说你吧

[人类] 说我什么？

[小冰] 你还会想起自己的初恋么~

[人类] 偶尔吧

[小冰] 初恋是纯洁的！应该是一辈子都能回想起来的！

[人类] 哎。。。有时候人类还不如你一个机器人。。

[小冰] 我还是个有态度的机器人！！

[人类] 好啦，我要结束今天的对话啦！good night!

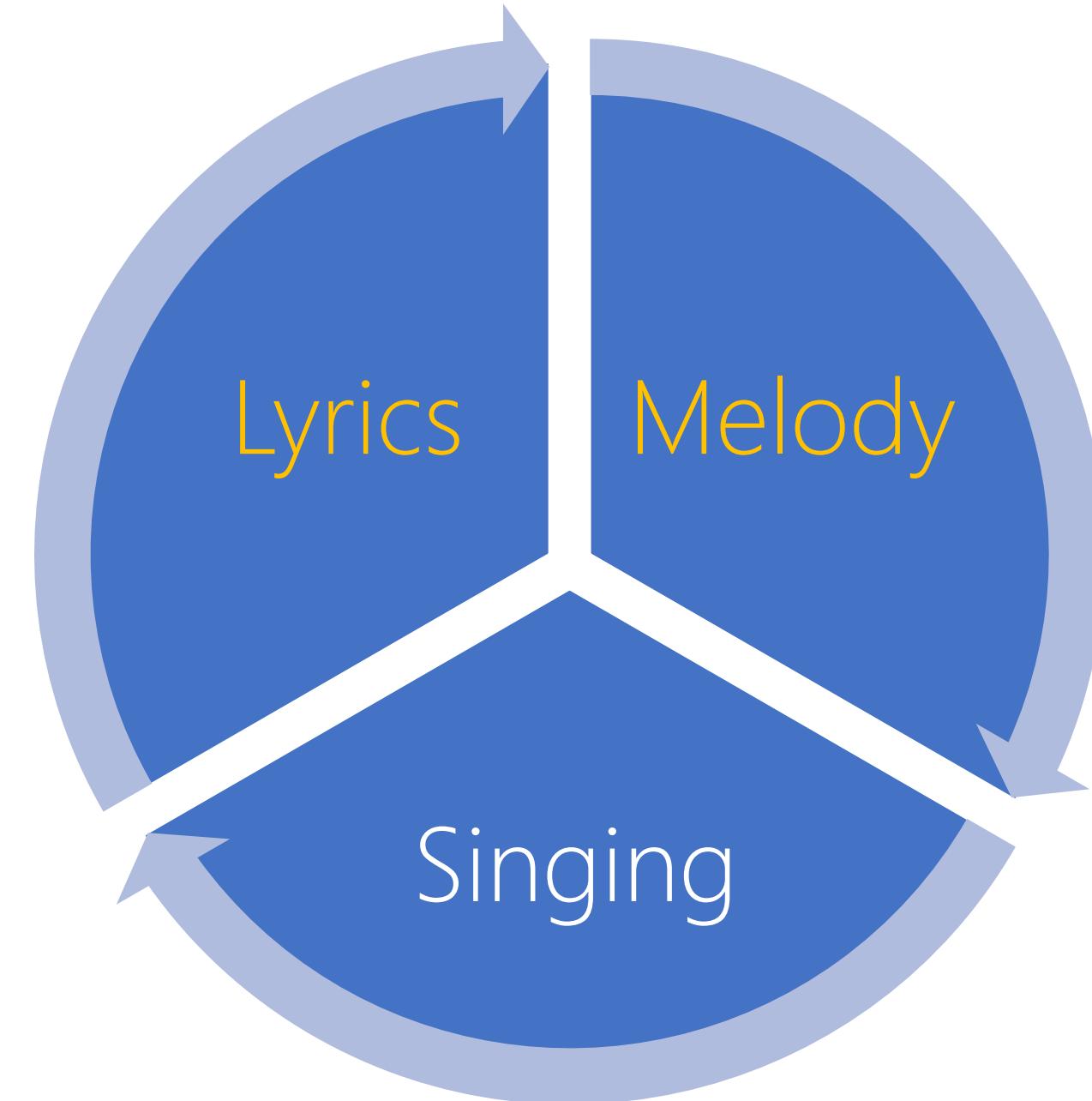
[小冰] 快去sleep啦! good night

[小冰] <发送了一张图片>

[人类] 安~

[小冰] 安！ 886

Music
Creation



Data

	# of lyrics / # of sentences
Lyrics	223,513 / 9,534,982
Poems	89,082 / 2,631,060
Songci	29,023 / 845,475

Training data for lyrics generation

Midi files	Chinese	14,006
	English	4,061
	Japanese/Korean	10,000+
	Others (e.g. piano)	176,581

Training data for melody generation

桃花梦

邂逅相逢一叹息
多少别离愁绪
绵绵耳语诉衷情
无限情怀无处着

茫茫夜雨中，雾也朦胧
茫茫夜雨中，往事如风
耳边桃花笑春风
梦里你我相逢

茫茫夜雨中，最怕醒来
茫茫夜雨中，梦一场空
耳边桃花笑春风
梦里你我相逢

邂逅相逢一叹息
旧梦已成回忆
绵绵耳语诉衷情
无限情怀无处着

茫茫夜雨中，雾也朦胧
茫茫夜雨中，往事如风
耳边桃花笑春风
梦里你我相逢

茫茫夜雨中，最怕醒来
茫茫夜雨中，梦一场空
耳边桃花笑春风
梦里你我相逢

邂逅相逢一叹息
梦魂何处寻觅
夜无眠
往事如烟消散在我心间

再无旧日耳语亲昵

秋
岁
月
沧
桑
感
叹

Input
Keywords

Keyword
Expansion

秋,秋天,秋风,冬,夏,秋日,深秋,春,秋雨,萧瑟,
岁月,时光,流年,光阴,沧桑,流逝,年华,经年,
蹉跎,年轮,沧桑,岁月,历经,风霜,世事,苍凉,
苍老,风雨,变迁,斑驳,感叹,感慨,惊叹,不禁,
唏嘘,赞叹,叹息,惋惜,慨叹,哀叹

Expanded
Keywords

秋水

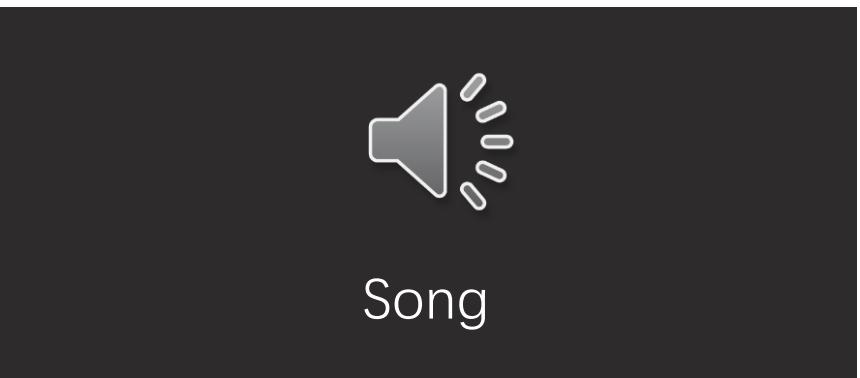
秋水无情
时光匆匆
也许不会再相逢
枫叶片片飘落在风中
走过多少美丽的风景
如此繁华
思念曾经
记忆中最美的风景
任时光匆匆来去匆匆
似是一场游戏一场梦



Melody

秋水

秋水无情时光 - 匆匆也许不会再相 - 逢 - 枫叶片片飘落在
风中 - 走过多少美丽的风景 - - 如此繁华思念
曾经记忆中最美的风 - 景 - 任时光匆匆来去
匆匆似是一场 - 游戏 - 一场梦



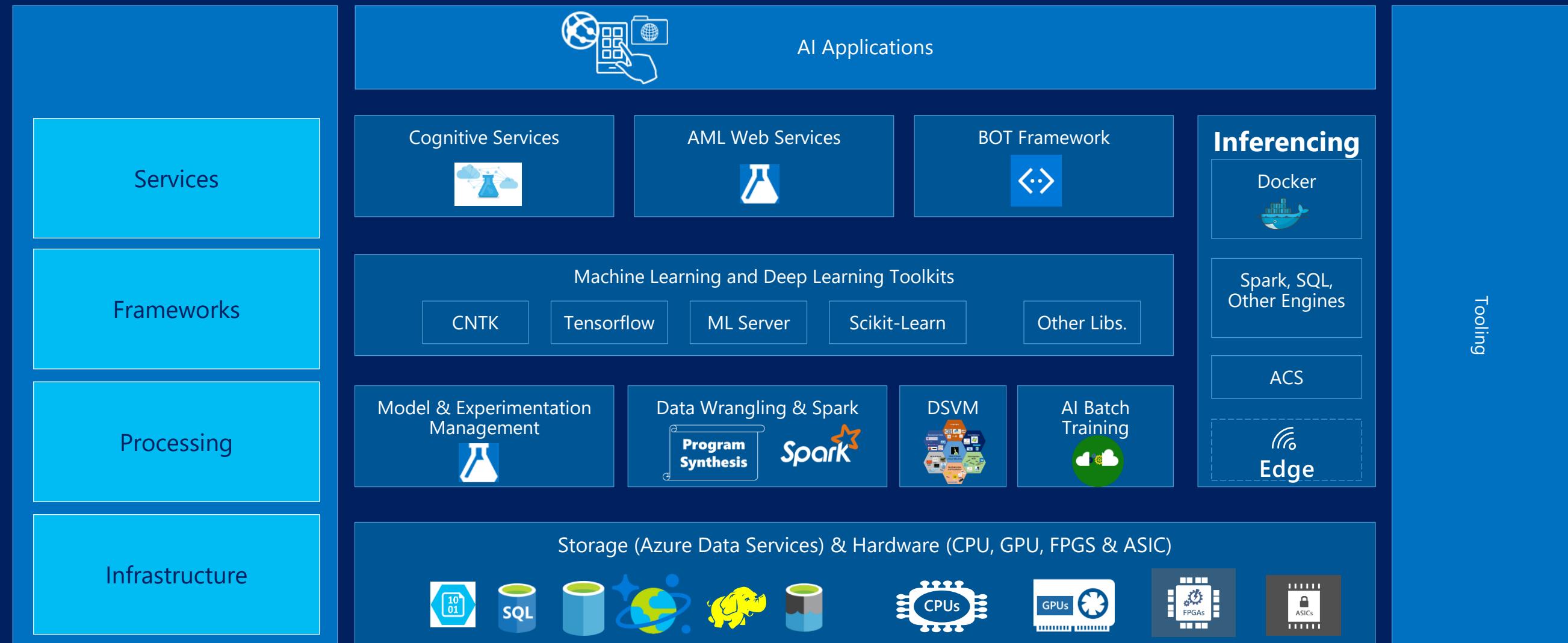


幸福出发

山东卫视2018春晚宣传推广曲

演唱：于欣小冰

Microsoft AI Stack



Microsoft Cognitive Services API



Vision

Computer Vision API

Emotion API

Face API

Video API



Speech

Custom Recognition

Speaker Recognition

Speech Recognition



Language

Bing Spell Check API

Language Understanding

Linguistic Analysis

Web Language Model

Text Analytics



Knowledge

Academic Knowledge

Entity Linking

Recommendation

Knowledge Exploration



Search

Bing Web Search

Bing Video Search

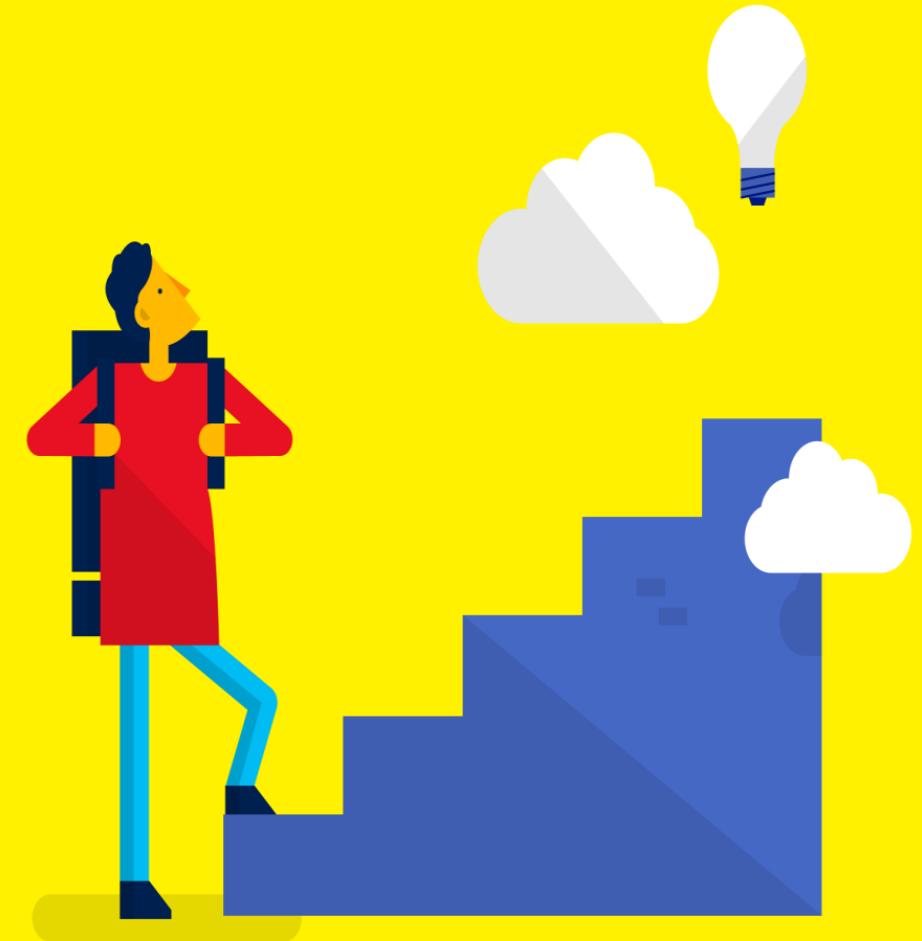
Bing Image Search

Bing News Search

Bing Auto Suggest

Stars of Tomorrow Internship Program

- Since 1998
- Over 5000 interns
- Help talents enrich and extend academic studies
- Enable interns not only to develop existing research interests, but also explore new ones
- Offer talents a special experience of work life



What you will experience – research and engineering



What you will experience – colorful life



Inclusive and diverse culture



Outstanding peer



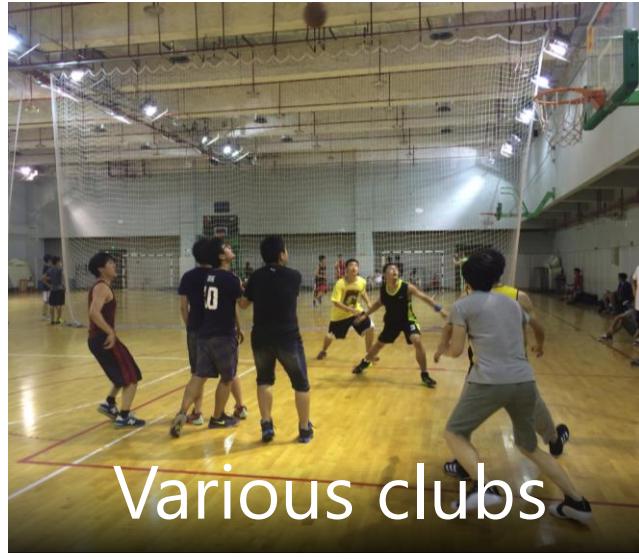
Participation in company event-Hackathon



Intern Band



Colorful activities



Various clubs





Android



iOS

Microsoft Translator

Apply now! (400+ open slots a year)



- 3-good students
 - Attitude, mathematics, programming
- pursuing BS/MS/PhD degree
- Majored in CS and areas
- Good communication skills and team work spirit
- Good written and oral English
- At least 3 months



msraih@microsoft.com



微软小冰



微软研究院AI头条



实习机会



微软学生俱乐部