

# Machine Learning for Speech Recognition

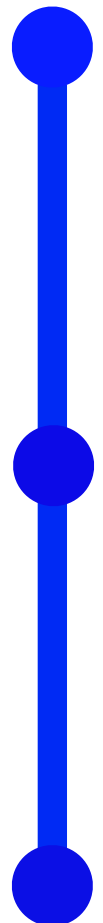
*by Alice Coucke, Head of Machine Learning Research*

*@alicecoucke*

*alice.coucke@sonos.com*

**snips** | **SONOS**

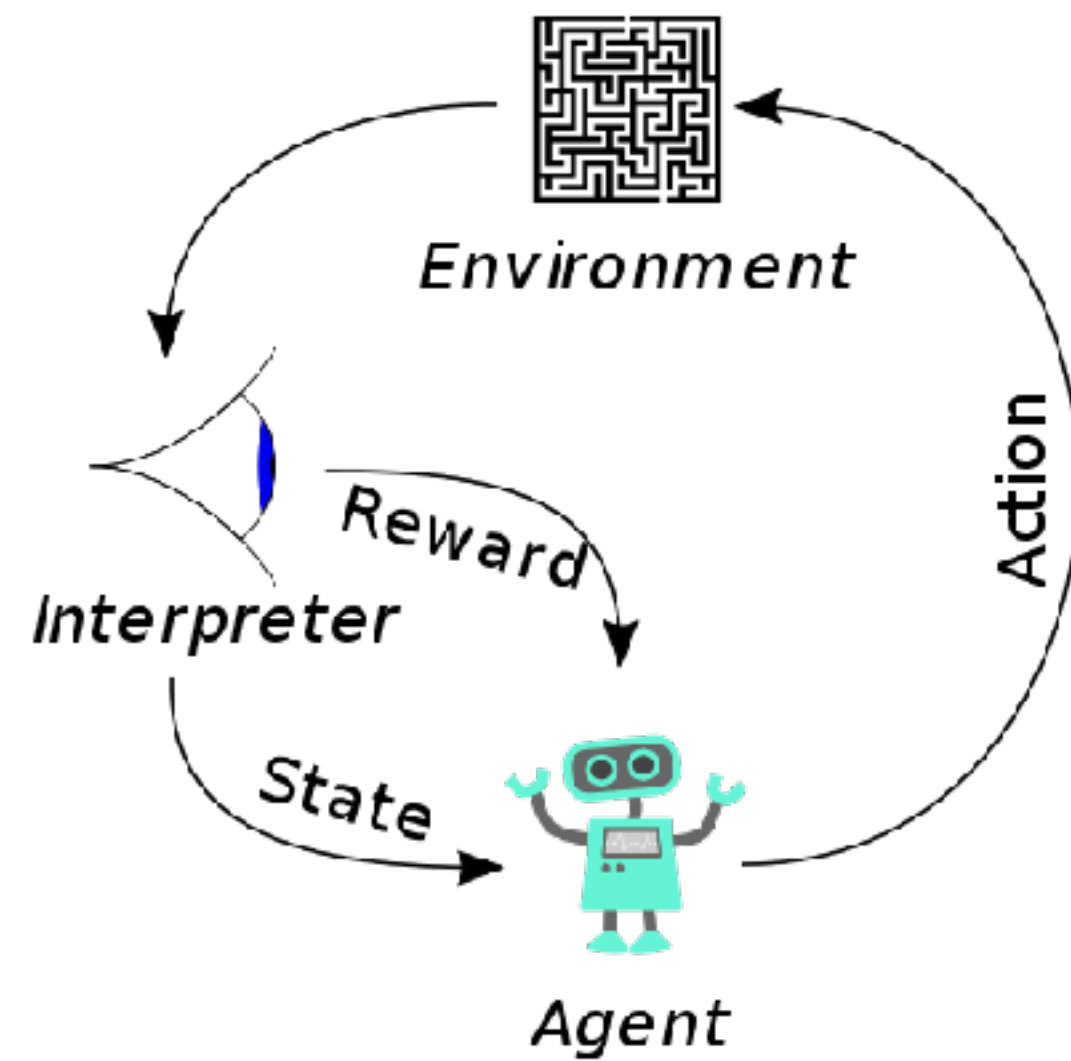
# Outline:

- 
- 1. Recent advances in machine learning**
  - 2. From physics to machine learning**
  - 3. Working at Snips (now Sonos)**

# Recent Advances in Applied Machine Learning

# Reinforcement learning

Learning goal-oriented behavior within simulated environments



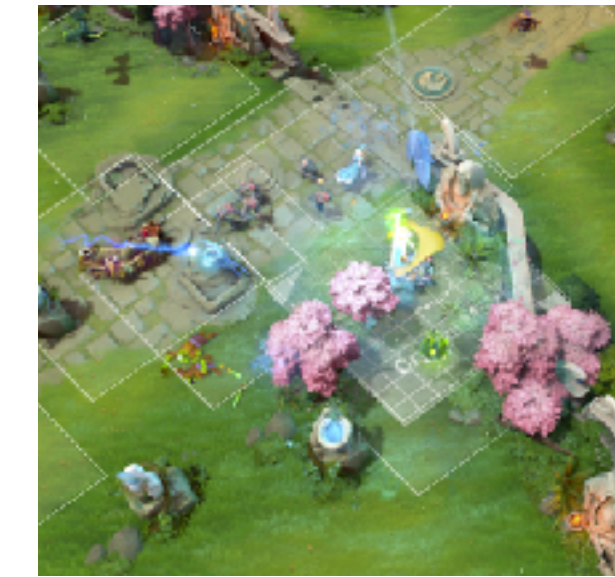
## Go

*AlphaGo (Deepmind, 2016)*



## Starcraft II

*AlphaStar (Deepmind)*



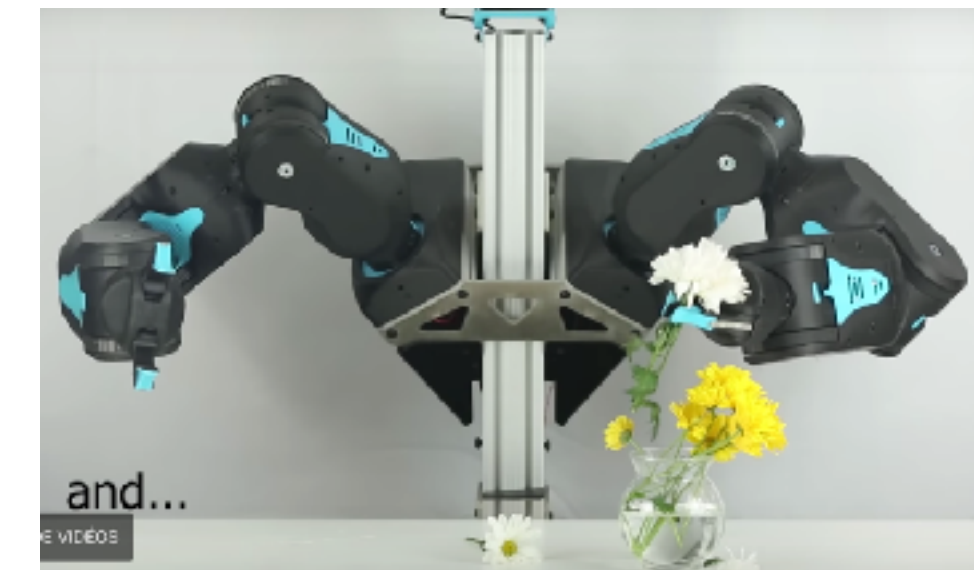
## Dota 2

*OpenAI Five (OpenAI)*



## Play-driven learning for robots

*(Google Brain)*

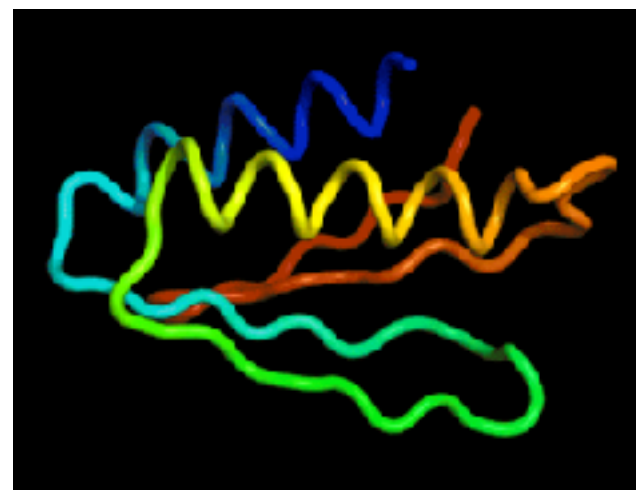


## Sim-to-real dexterity learning

*Project BLUE (UC Berkeley)*

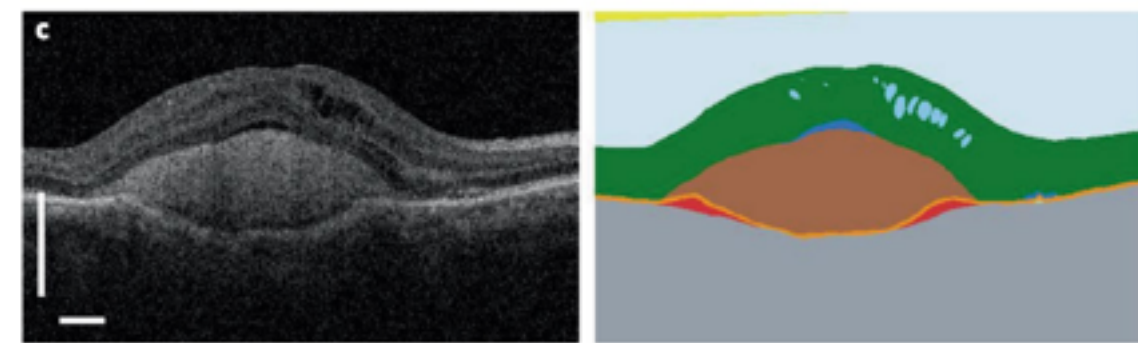
# Machine Learning for Life Sciences

Deep learning applied to biology and medicine



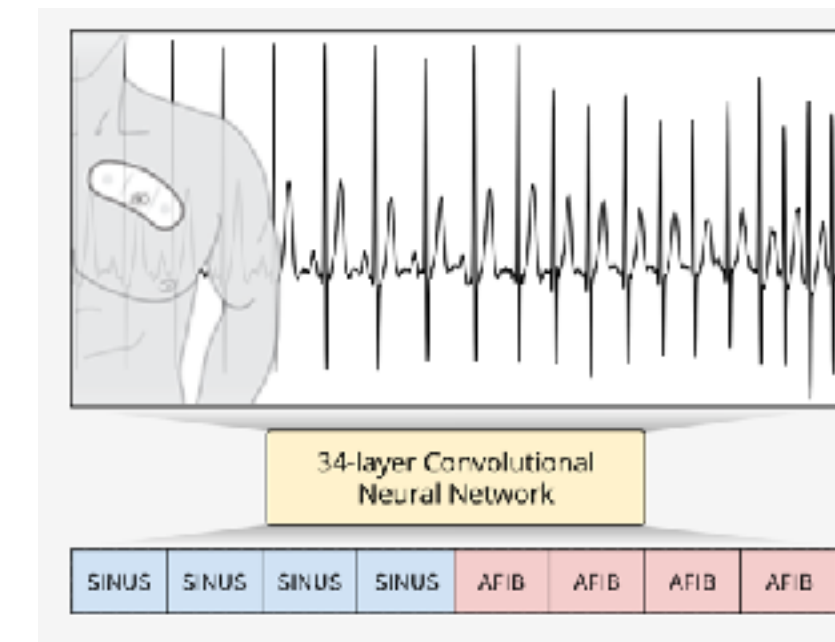
## Protein folding & structure prediction

*AlphaFold (Deepmind)*



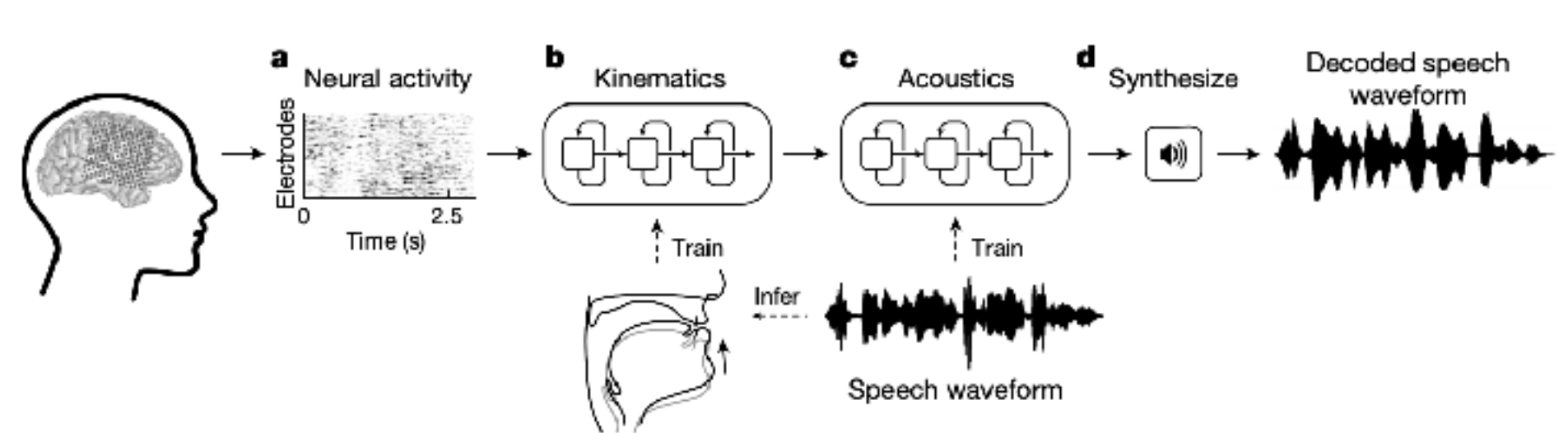
## Eye disease diagnosis

*(NHS, UCL, Deepmind)*



## Cardiac arrhythmia prediction from ECGs

*(Stanford)*



## Reconstruct speech from neural activity

*(UCSF)*



## Limb control restoration

*(Batelle, Ohio State Univ)*

# Computer vision

High-level understanding of digital images or videos



(a) Ground Truth: Stacking 4 coins.  
(b) Model output: Piling coins up.



(c) Ground Truth: Lifting up one end of flower pot, then letting it drop down.  
(d) Model output: Lifting up one end of bucket, then letting it drop down.

« Common sense » understanding  
of actions in videos

*(TwentyBn, DeepMind, MIT, IBM...)*



**GANs for image generation**

*(Heriot Watt Univ, DeepMind)*



**GANs for artificial video dubbing**

*(Synthesia)*



**GAN for full body synthesis**

*(DataGrid)*

# From physics to machine learning and back

A surge of interest from the physics community



## NeurIPS 2019: workshop on « machine learning and the physical sciences »

- Application of machine and deep learning to physical sciences
- Generative models
- Likelihood-free inference
- Variational inference
- Simulation-based models
- Implicit models
- Probabilistic models
- Model interpretability
- Approximate Bayesian computation
- Strategies for incorporating prior scientific knowledge into machine learning algorithms
- Experimental design
- Any other area related to the subject of the workshop

### The jamming transition as a paradigm to understand the loss landscape of deep neural networks

Mario Geiger,<sup>1,\*</sup> Stefano Spigler,<sup>1,\*</sup> Stéphane d'Ascoli,<sup>2,3</sup> Levent Sagun,<sup>2,1</sup> Marco Baity-Jesi,<sup>4</sup> Giulio Biroli,<sup>2,3</sup> and Matthieu Wyart<sup>1</sup>

<sup>1</sup>*Institute of Physics, EPFL, CH-1015 Lausanne, Switzerland*

<sup>2</sup>*Institut de Physique Théorique, Université Paris-Saclay, CEA, CNRS, F-91191 Gif-sur-Yvette, France*

<sup>3</sup>*Laboratoire de Physique Statistique, École Normale Supérieure, PSL Research University, F-75005 Paris, France*

<sup>4</sup>*Department of Chemistry, Columbia University, 10027 New York, USA*

(Dated: June 18, 2019)

### Modelling the influence of data structure on learning in neural networks

Sebastian Goldt<sup>1</sup>, Marc Mézard<sup>2</sup>,  
Florent Krzakala<sup>2</sup> and Lenka Zdeborová<sup>1</sup>

<sup>1</sup>*Institut de Physique Théorique, CNRS, CEA, Université Paris-Saclay, France*

<sup>2</sup>*Laboratoire de Physique de l'École Normale Supérieure, Université PSL, CNRS, Sorbonne Université, Université Paris-Diderot, Sorbonne Paris Cité, Paris, France*

### Capacity-resolution trade-off in the optimal learning of multiple low-dimensional manifolds by attractor neural networks

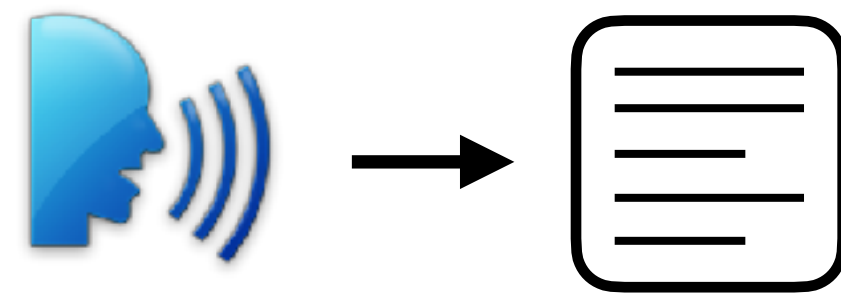
Aldo Battista and Rémi Monasson

*Laboratory of Physics of the Ecole Normale Supérieure,  
CNRS UMR 8028 & PSL Research, Paris, France*

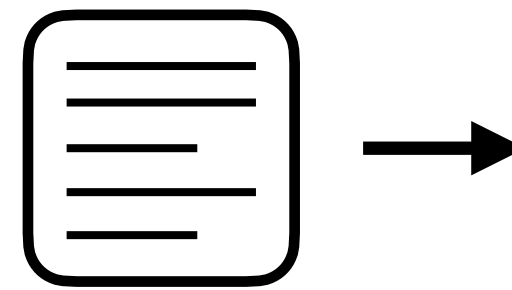
(Dated: October 15, 2019)

# Speech and language

Understand and analyze human speech

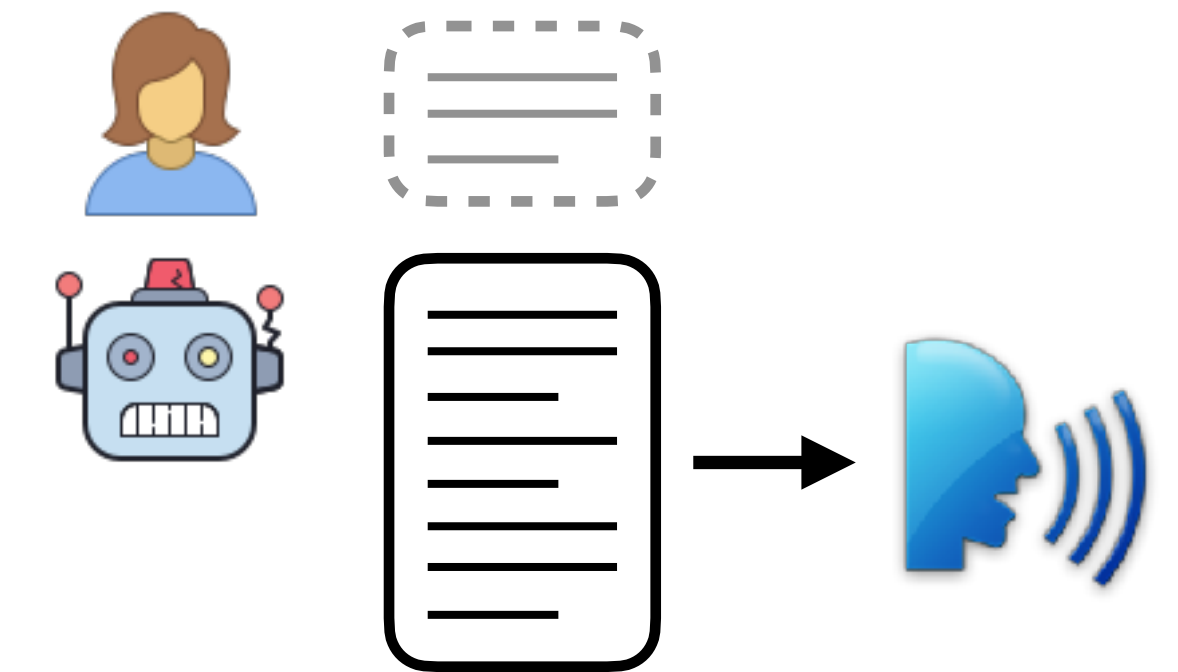


**Speech transcription**  
*Human Parity (Microsoft)*



```
{  
  intent: FindWeather,  
  entities: {  
    datetime: 11/28/2019,  
    location: Paris  
  }  
}
```

**Spoken language understanding**  
*(Super)GLUE benchmarks (Google, Facebook, IBM, Stanford ...)*

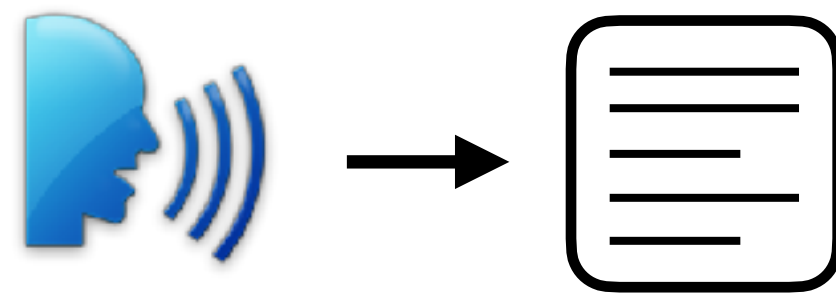


**Text & speech generation**  
*GPT-2 (Open AI)  
Bert (Google)  
XLNet (CMU) ...*

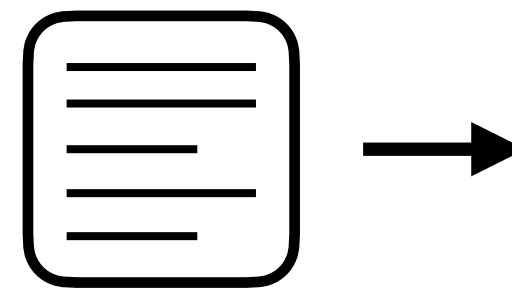


# Speech and language

Understand and analyze human speech

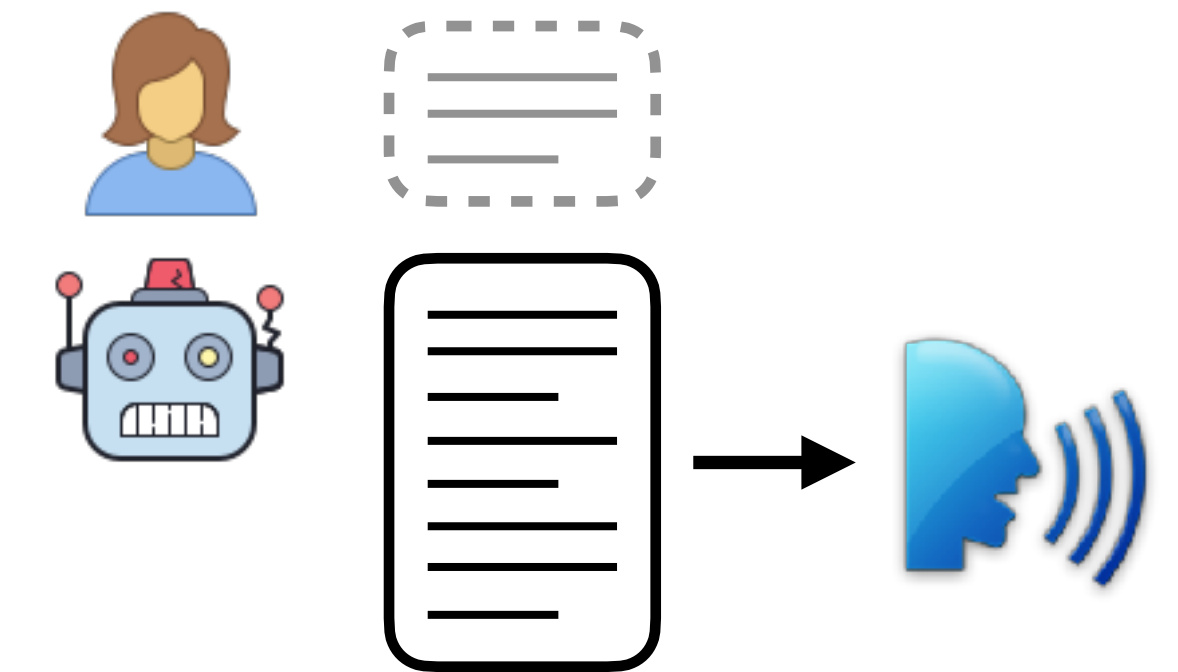
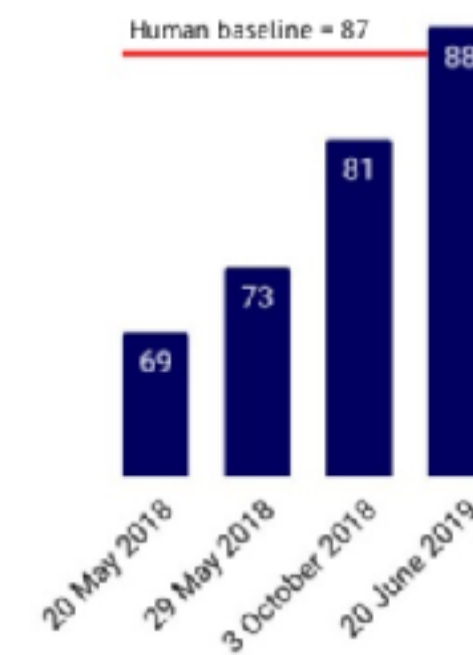


**Speech transcription**  
*Human Parity (Microsoft)*



```
{  
  intent: FindWeather,  
  entities: {  
    datetime: 11/28/2019,  
    location: Paris  
  }  
}
```

**Spoken language understanding**  
*(Super)GLUE benchmarks (Google, Facebook, IBM, Stanford ...)*



**Text & speech generation**  
*GPT-2 (Open AI)  
Bert (Google)  
XLNet (CMU) ...*



**Neural machine translation**  
*Unsupervised MT (Facebook)*



**Voice activity detection**  
*Detect speech from audio*



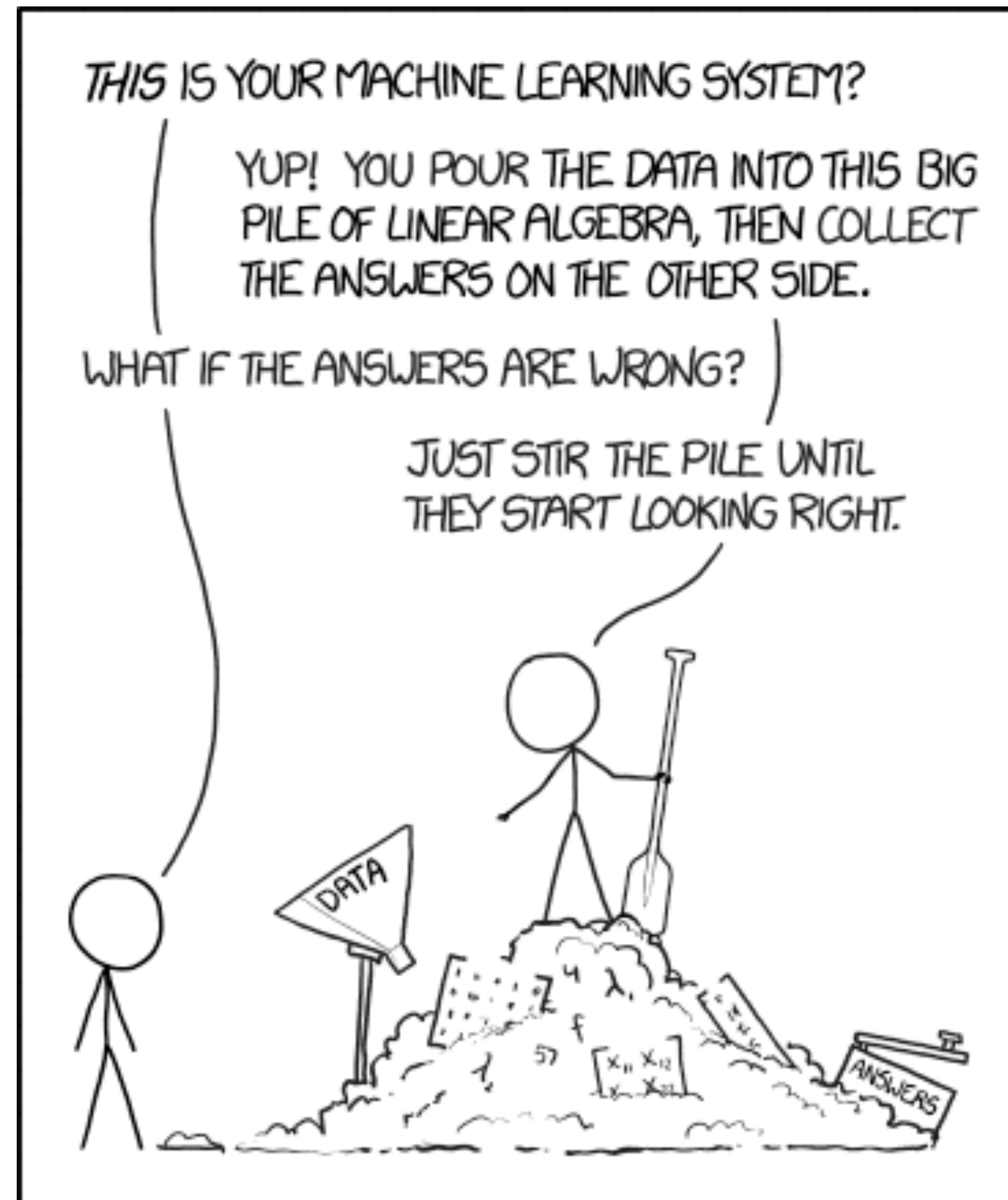
**Speaker identification**  
*Recognize unique speakers*



**Sentiment analysis**  
*Detect emotions in text and speech*

# Fairness, ethics, and explainability

We, as scientists, have a say in the future of AI



# From physics to machine learning

# **My background**

From physics to machine learning

# My background

From physics to machine learning



**2012:** M2 ICFP Theoretical physics

**2013-2016:** PhD in statistical physics @LPTENS

# My background

From physics to machine learning



**2012:** M2 ICFP Theoretical physics

**2013-2016:** PhD in statistical physics @LPTENS



# My background

From physics to machine learning



**2012:** M2 ICFP Theoretical physics

**2013-2016:** PhD in statistical physics @LPTENS



**Feb 2017:** senior ML scientist @ Snips

**2019:** director of ML research @ Snips

**Today:** head of ML research @ Sonos, Inc.

# **A few takeaways**

(please go ask other people too)



# A few takeaways

(please go ask other people too)

- PhD? Postdoc?
- Working at a startup company
- Physicists and machine learning

# A few takeaways

(please go ask other people too)

- PhD? Postdoc?
- Working at a startup company
- Physicists and machine learning

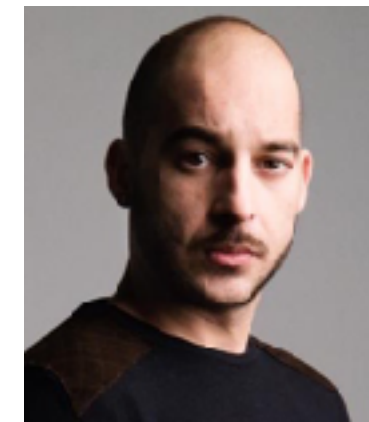
## Physicists @ Snips



**A. C.**



**Raffaele Tavarone**  
*Sr ML Scientist*  
*Acoustics team*



**Francesco Caltagirone**  
*Sr ML Scientist*  
*Tech Lead*  
*Language team*



**Alaa Saade**  
*Sr ML Scientist*  
*Now: DeepMind*



**Stéphane d'Ascoli**  
*ML research intern*  
*Now: PhD ENS & FAIR*

# A few takeaways

(please go ask other people too)

- PhD? Postdoc?
- Working at a startup company
- Physicists and machine learning

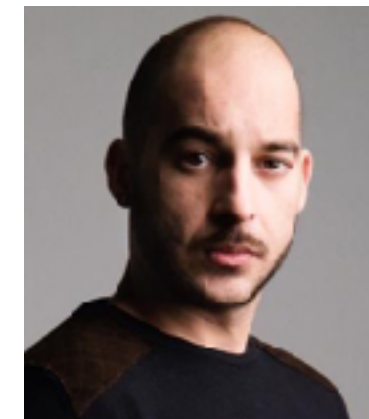
## Physicists @ Snips



**A. C.**



**Raffaele Tavarone**  
*Sr ML Scientist*  
*Acoustics team*



**Francesco Caltagirone**  
*Sr ML Scientist*  
*Tech Lead*  
*Language team*



**Alaa Saade**  
*Sr ML Scientist*  
*Now: DeepMind*



**Stéphane d'Ascoli**  
*ML research intern*  
*Now: PhD ENS & FAIR*



Uncharted technologies

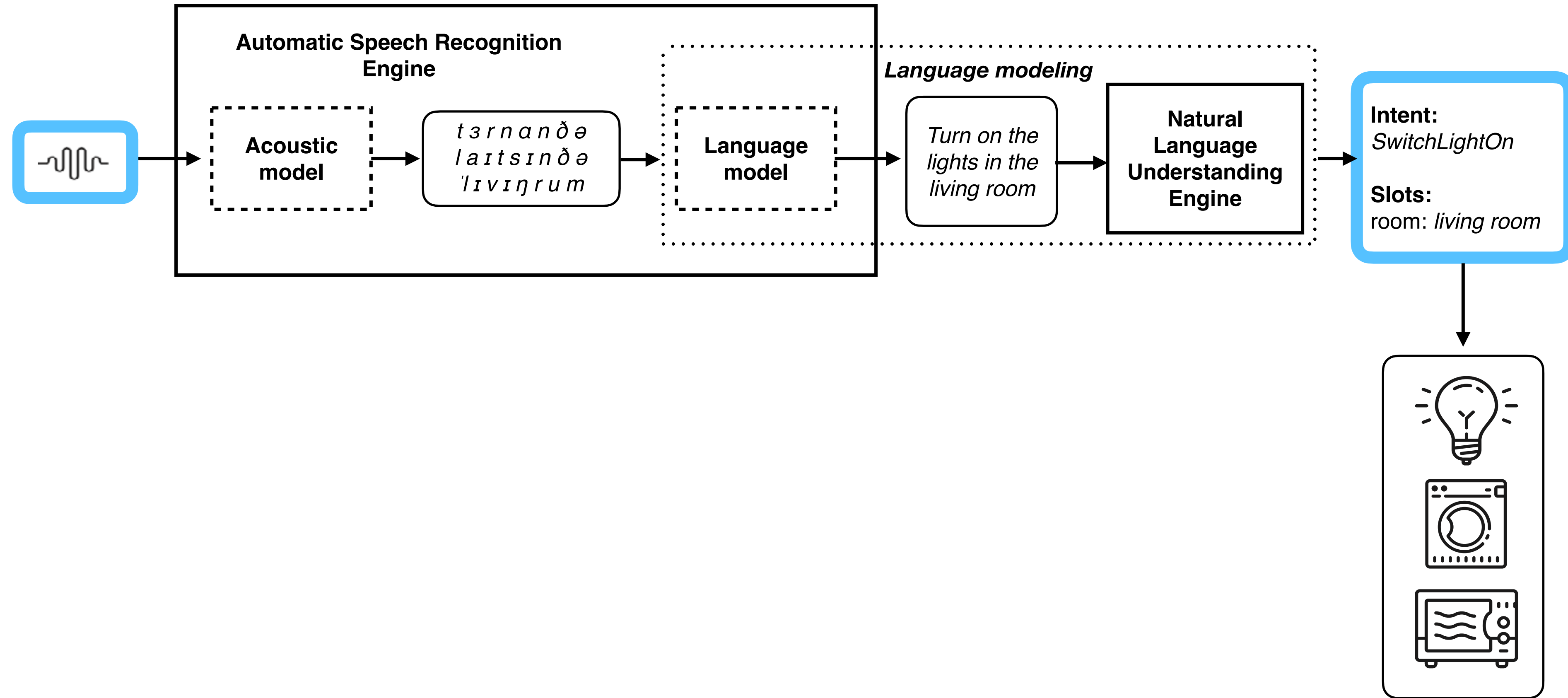


# Machine Learning at Snips (now Sonos)

**snips** | **SONOS**

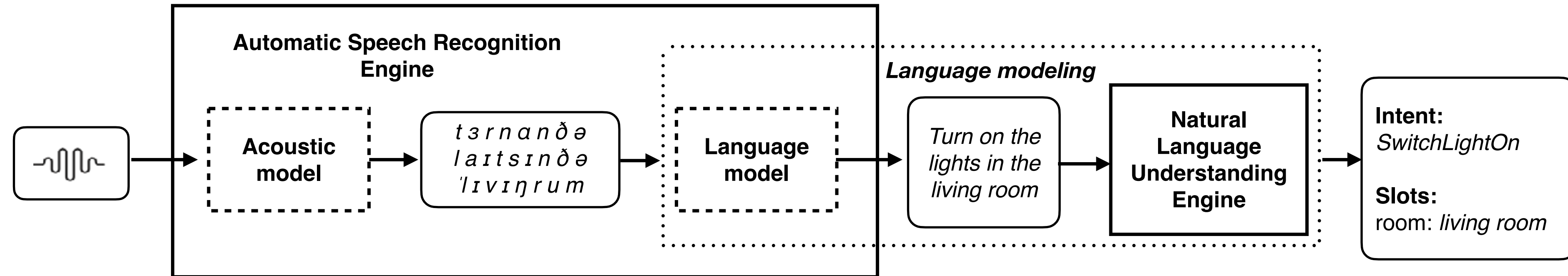
# Spoken Language Understanding

From speech to meaning for voice assistants



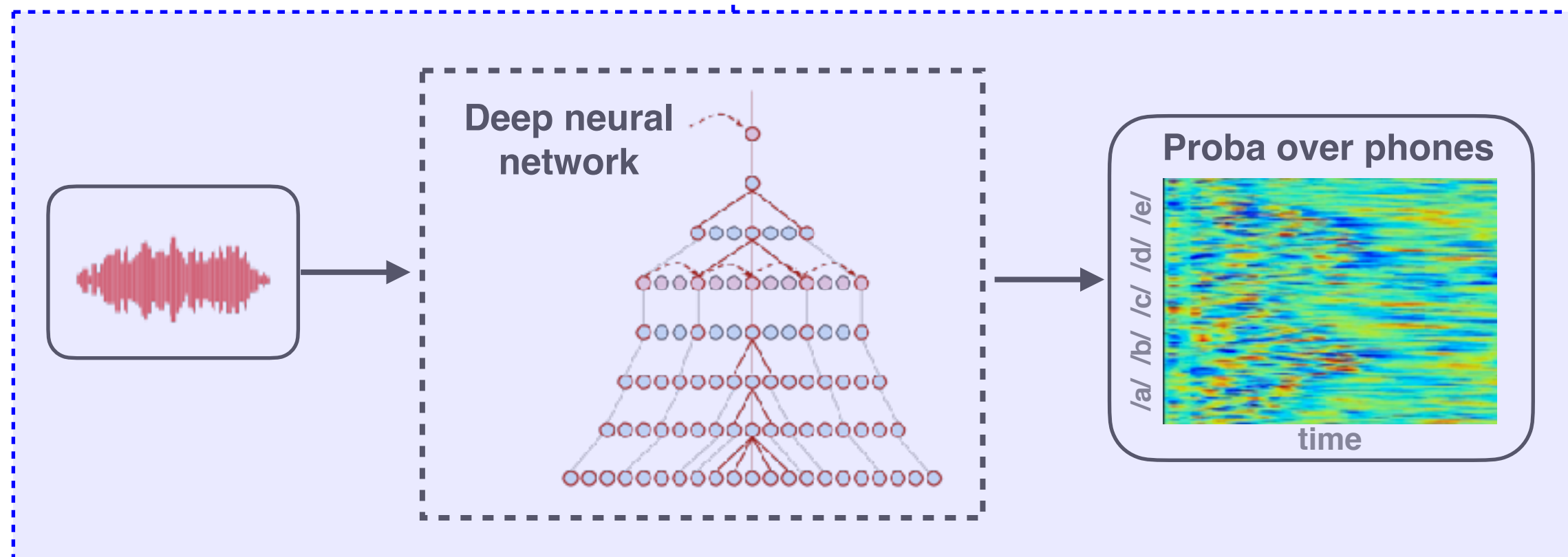
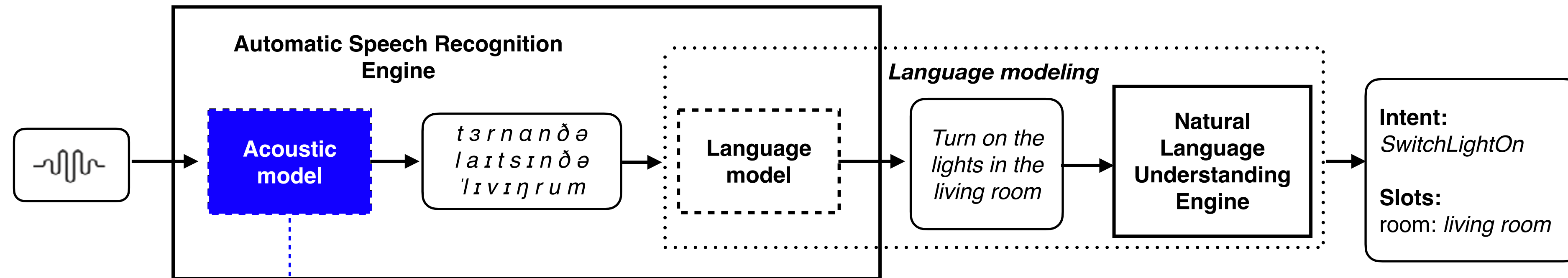
# Spoken Language Understanding

From speech to meaning for voice assistants



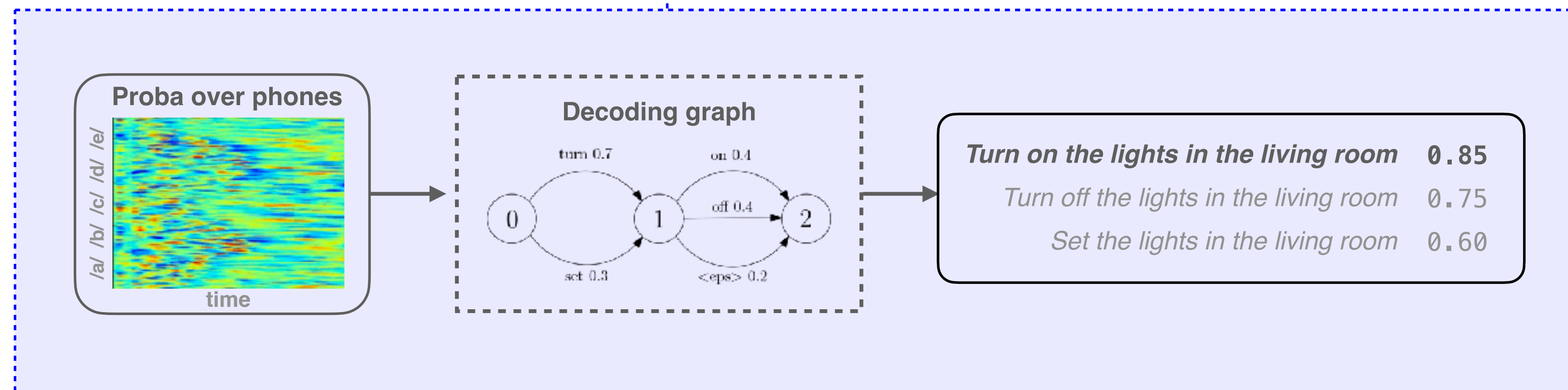
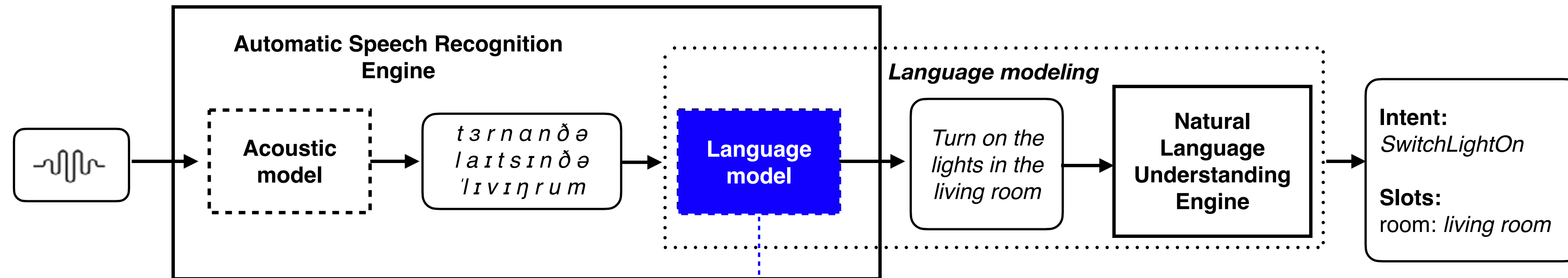
# Spoken Language Understanding

From speech to meaning for voice assistants



# Spoken Language Understanding

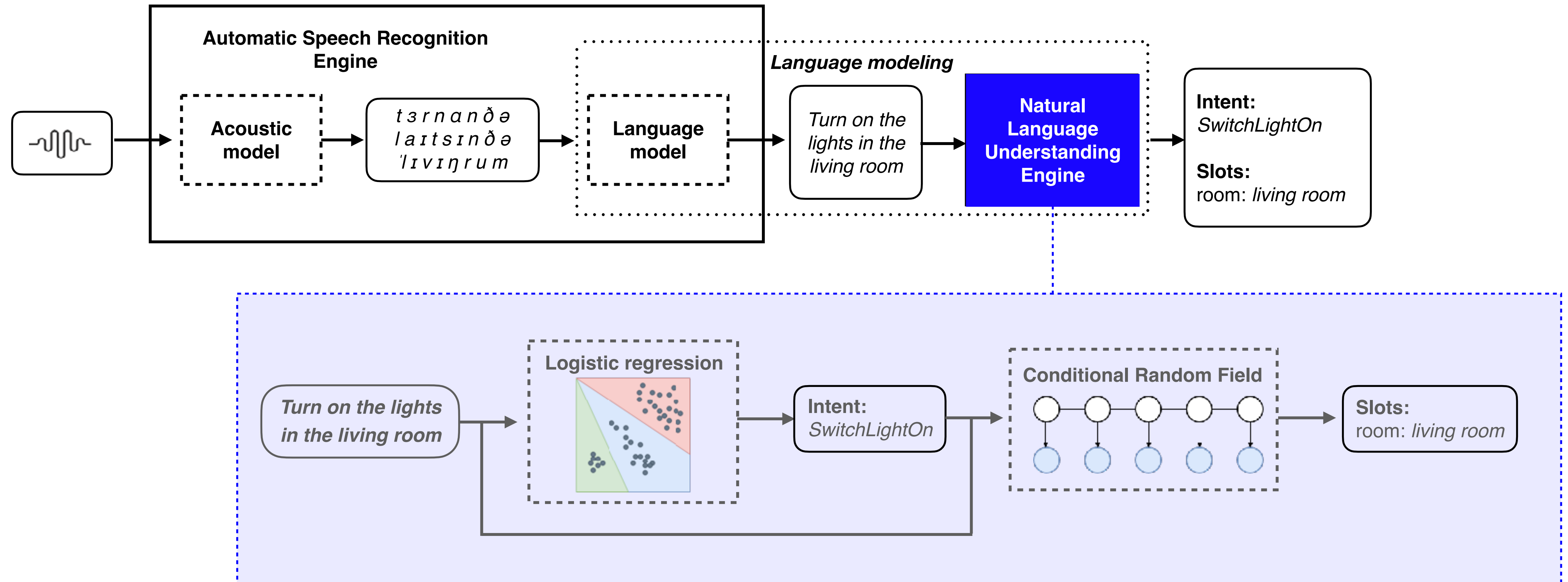
From speech to meaning for voice assistants





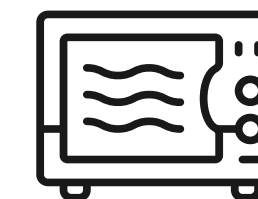
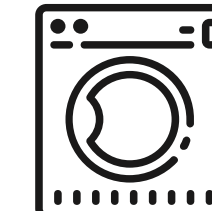
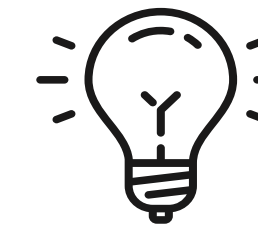
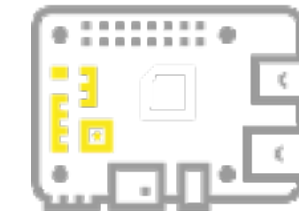
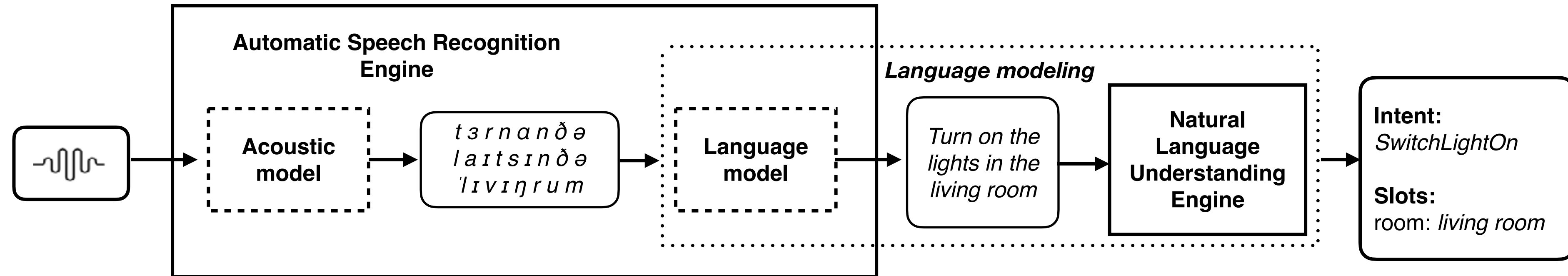
# Spoken Language Understanding

From speech to meaning for voice assistants



# Spoken Language Understanding

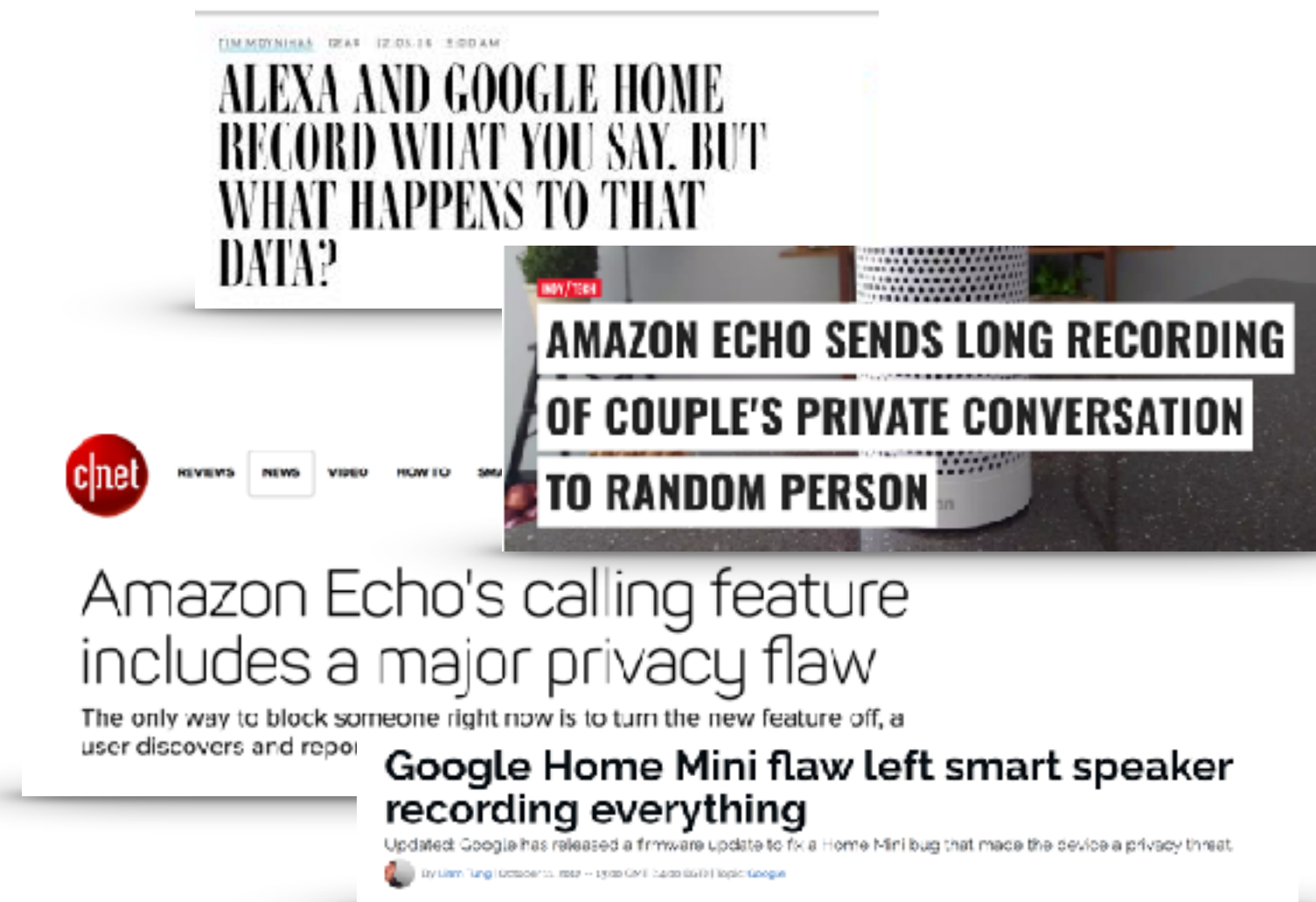
Offline & on device



# Our approach

Our own voice in a vast ecosystem

**Privacy by design**



**Resource constrained ML:  
small data & hardware**



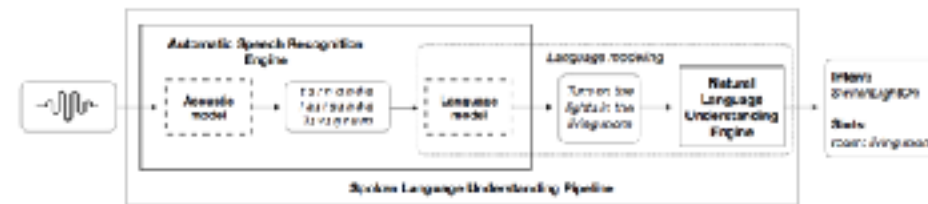
- ✓ A new **popular trend** in the ML community (low resource ML, transfer learning, miniaturization, etc)
- ✓ Numerous conferences and workshops on the topic
- ✓ Towards a **safer, greener and more private** conversational AI

# Research activity

## Publishing research in industry

**Snips voice platform:**  
an embedded spoken language  
understanding system for  
private-by-design voice  
interfaces

*arXiv*



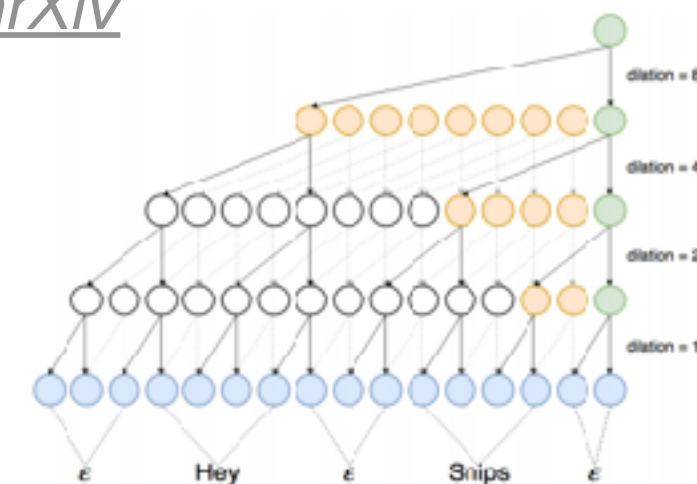
ICML 2018  
workshop PiMLAI



Cited 35 times  
to date

**Efficient keyword spotting**  
using dilated convolutions  
and gating

*arXiv*

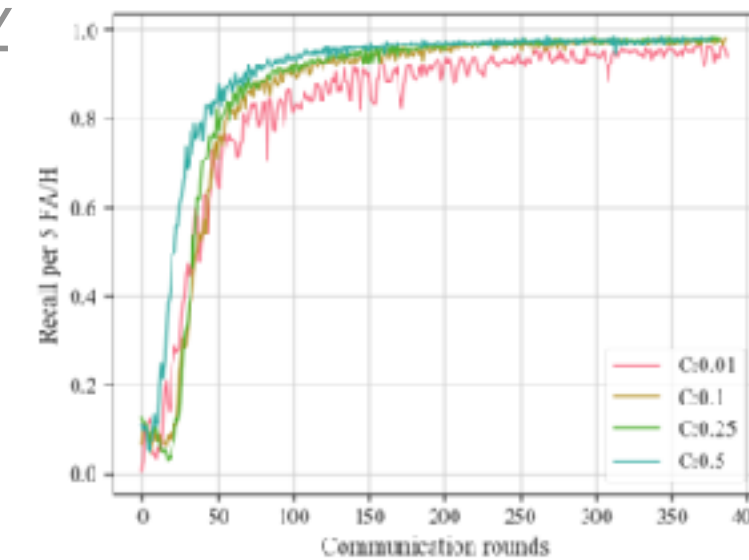


ICASSP 2019  
Main track



**Federated learning for  
keyword spotting**

*arXiv*



ICASSP 2019  
Main track



**Spoken language  
understanding on the edge**

*arXiv*

Quantity	Close field		Far field	
	Snips	Google	Snips	Google
Intent classification (F1)	0.92	0.89	0.84	0.86
Perfect parsing (%)	0.84	0.79	0.72	0.73

**Table 3.** End-to-end generalization performance on the “SmartLights” assistant: comparison with Google’s Dialogflow cloud service on a 5-fold cross-validation experiment.

NeurIPS 2019  
Workshop EMC2



✓ Publish **open & reproducible** benchmarks:

- ▶ ~200 access granted to researchers to our **open speech datasets**
- ▶ Snips dataset for NLU is the **new academic standard**

**Thank you for your attention**  
*Questions?*

*@alicecoucke*

*alice.coucke@sonos.com*

**snips** | **SONOS**

