Building Systems that Talk to Us

DESH RAJ
Johns Hopkins University

https://desh2608.github.io

LMUCON'20

A bit about me…

B.Tech.
Computer Science
2013-17

Research Intern
Summer 2015

Software Engineering Intern
Summer 2016

Research engineer
Advanced Technology Lab
June 2017 - June 2018

Ph.D. student
Computer Science
since Fall 2018

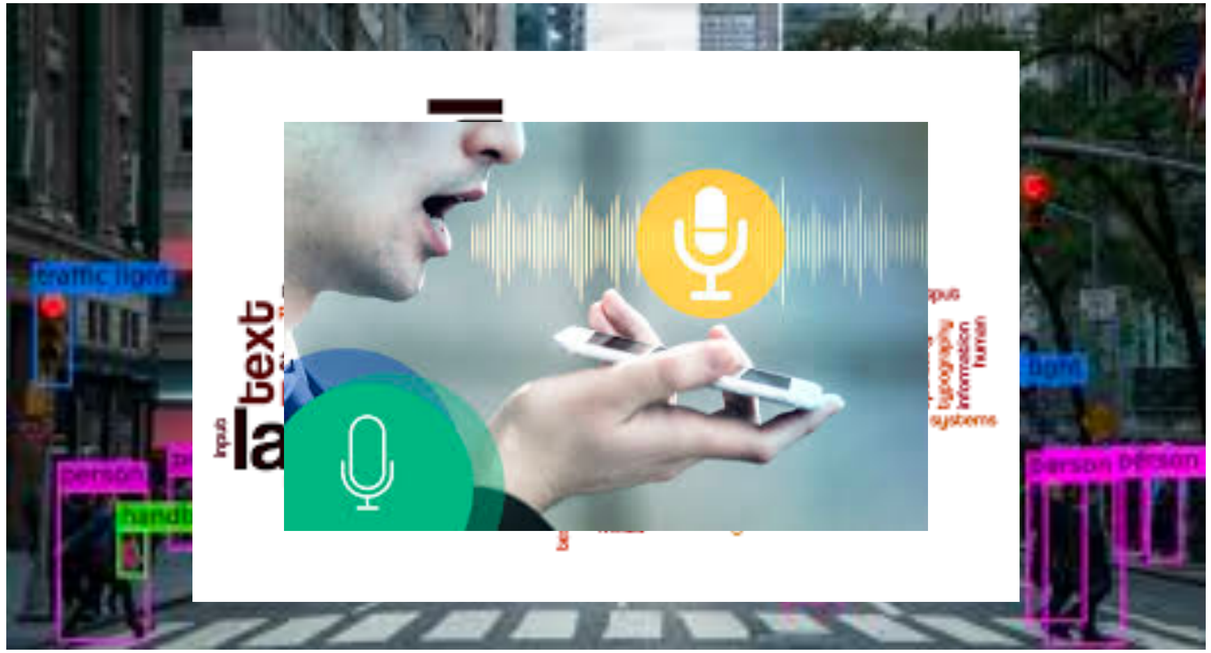
A bit more about me…

Did some computer vision stuff at first

A bit more about me…

Worked on NLP for my Undergrad Thesis

A bit more about me…



Now doing a Ph.D. in speech recognition

Conversational systems are all around us…

Personal assistants

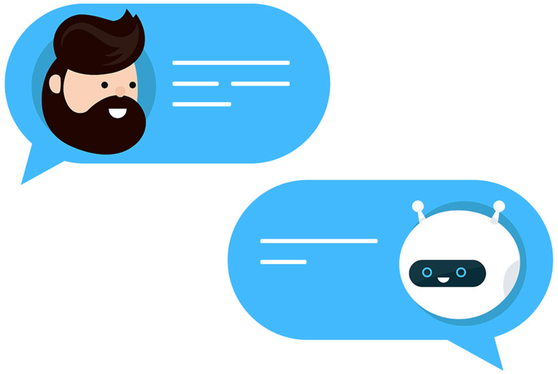
Apple Siri

Ok Google!

Microsoft Cortana

Smart home devices

Conversational systems are all around us…

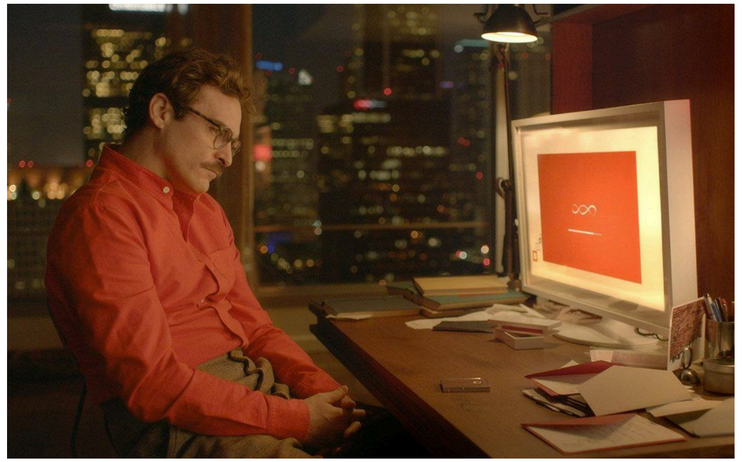


Automated customer service

In-car voice assistant

…and even in movies

Iron Man



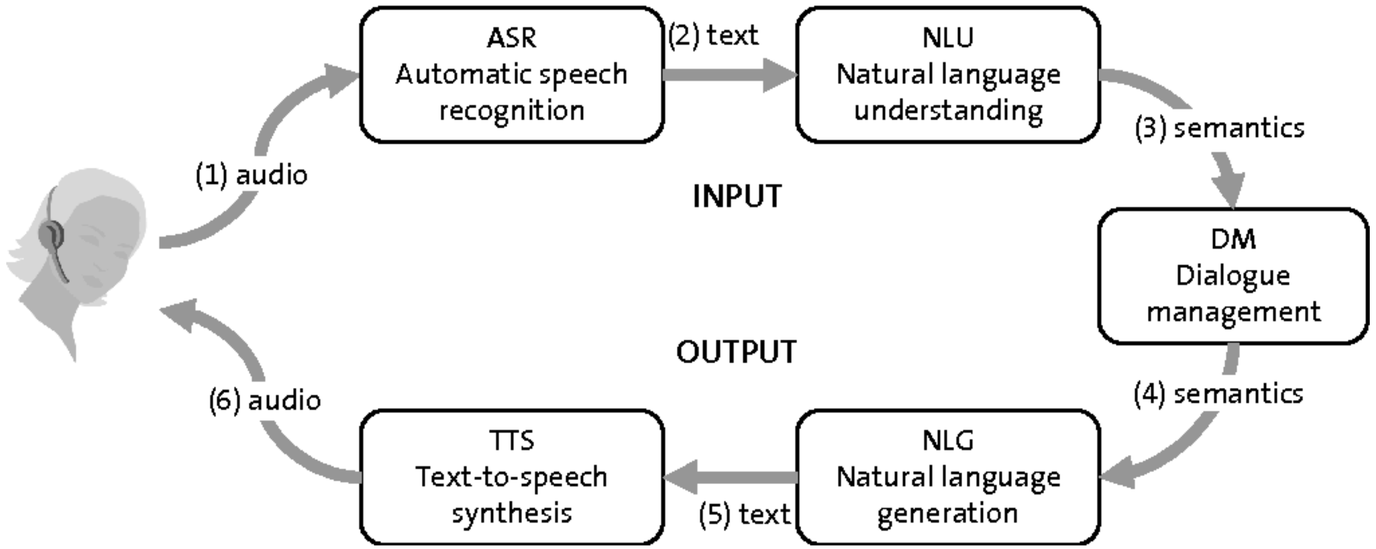
HER

Black Mirror

How are these systems built?

- **Speech processing:** recognize who said what, speak back to user
- **Natural language processing (NLP):** get meaning from utterance, generate response to user
- **Computer vision:** get context from environment

A typical dialog system pipeline



"Error Handling in Spoken Dialogue Systems," Gabriel Skantze. Doctoral Thesis.

And not just for dialog systems!

Speech

Meeting transcription

NLP

Speech recognition

Speaker diarization

Speaker tracking

Vision

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

NLP

Speech recognition

Speech activity detection

Vision

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Speech recognition
Speech synthesis

NLP

Vision

And not just for dialog systems!

Speech

NLP

Vision

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Speaker recognition

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

NLP

Speaker/emotion recognition

Time series analysis

Vision

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

Removing noise from speech

NLP

Speech enhancement

Vision

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

Removing noise from speech

NLP

Auto-complete

Language modeling

Vision

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

Removing noise from speech

NLP

Auto-complete

Language translation

Vision

Neural machine translation

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

Removing noise from speech

NLP

Auto-complete

Language translation

Semantic search

Semantic parsing

Syntactic parsing

Vision

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

Removing noise from speech

NLP

Auto-complete

Language translation

Semantic search

Document summarization

Vision

Automatic summarization

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

Removing noise from speech

NLP

Auto-complete

Language translation

Semantic search

Document summarization

Vision

Autonomous vehicles

Object detection

Scene segmentation

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

Removing noise from speech

NLP

Auto-complete

Language translation

Semantic search

Document summarization

Vision

Autonomous vehicles

Image summary generation

Image captioning

Language modeling

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

Removing noise from speech

NLP

Auto-complete

Language translation

Semantic search

Document summarization

Vision

Autonomous vehicles

Image summary generation

Image/video search

Image classification

And not just for dialog systems!

Speech

Meeting transcription

Automatic subtitle generation

Assisting hearing/speaking impaired people

Voice-based authentication

Diagnosing diseases such as Parkinson’s

Removing noise from speech

NLP

Auto-complete

Language translation

Semantic search

Document summarization

Vision

Autonomous vehicles

Image summary generation

Image/video search

Diagnosing illnesses e.g., find tumors

Image classification/segmentation

It is an exciting time for AI research

Speech

NLP

Vision

It is an exciting time for AI research

Speech

NLP

Vision

Earlier:

- All these tasks had different (often complex) strategies
- Required a great deal of handcrafted feature engineering

It is an exciting time for AI research

Speech NLP Vision

Earlier:

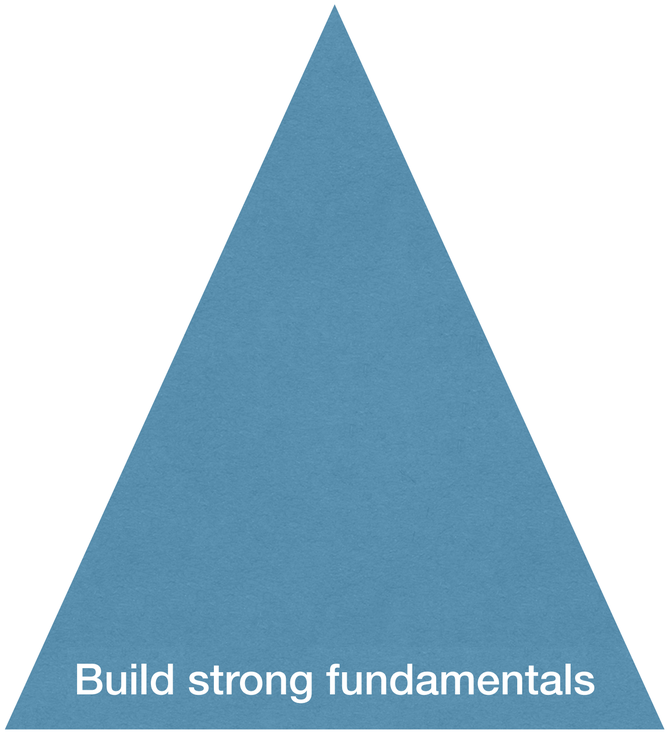
- All these tasks had different (often complex) strategies
- Required a great deal of handcrafted feature engineering

Now:

DEEP LEARNING + more data + more computational resources

How can I get started with NLP/Speech research?

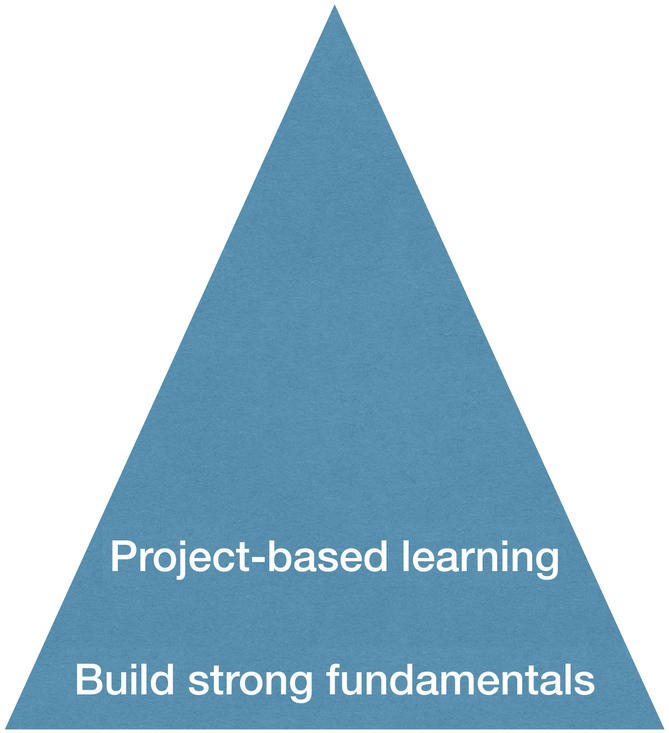
The Pyramid approach



- Linear algebra, probability
- Basic programming/scripting
- Core AI/ML -> searching, first order logic, statistical ML

How can I get started with NLP/Speech research?

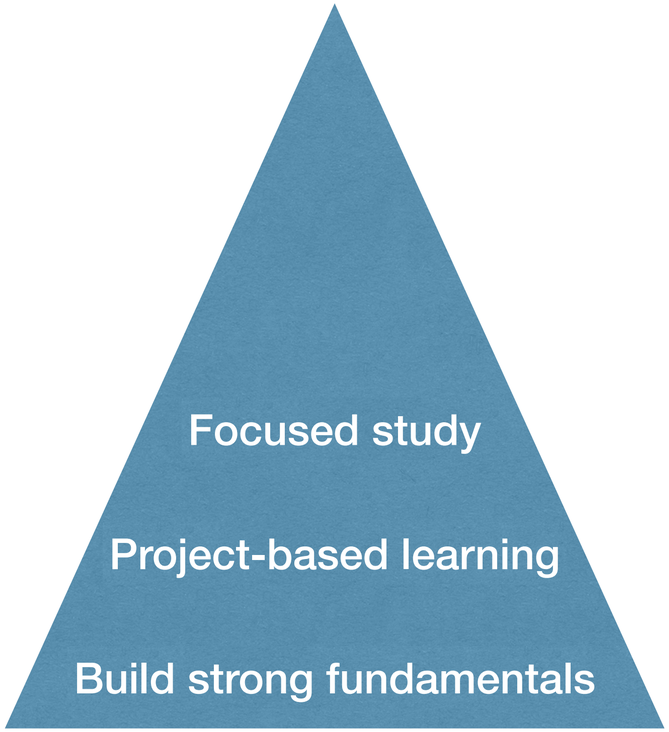
The Pyramid approach



- Easy **course based projects** (or online resources)
- Focus on learning tools like **Tensorflow, PyTorch**, etc.

How can I get started with NLP/Speech research?

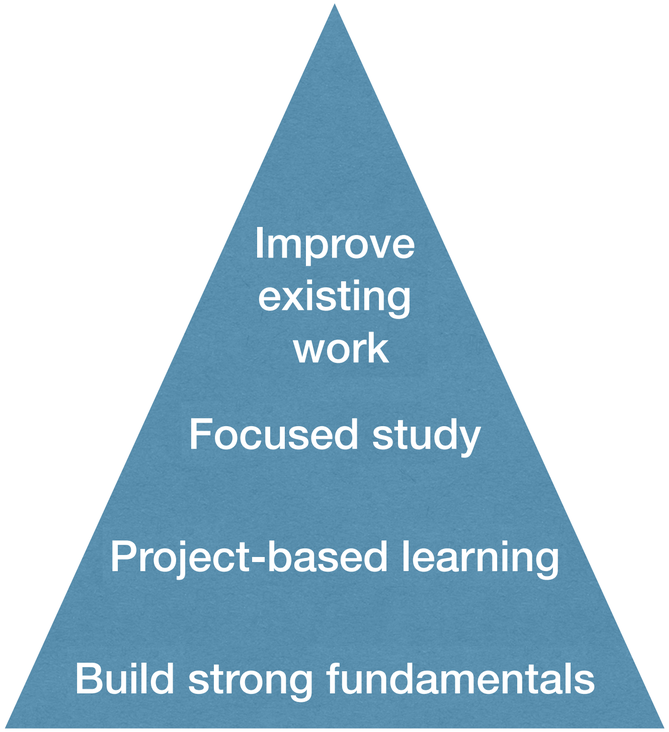
The Pyramid approach



- Start focusing on **one modality**/task, e.g., I started with relation extraction in my undergrad
- **Read papers** and adapt core ideas for your task
- Think about ways to extend these ideas

How can I get started with NLP/Speech research?

The Pyramid approach



- For ideas, look at **related tasks in other fields.** For example, the Transformer model originally proposed for machine translation is now also used in vision and speech recognition.
- Build a strong research profile with **open-source projects**

But there is so much out there!

ML/Deep Learning Libraries

Source: https://towardsdatascience.com/best-python-libraries-for-machine-learning-and-deep-learning-b0bd40c7e8c

Speech Recognition Toolkits

But there is so much out there!

- Learning ML/DL/AI is not equal to learning tools
- The project you work on should guide the tool you use, and not vice-versa
- **Diversifying tools/projects** often makes for a strong profile -> better candidate for industrial jobs

But there is so much out there!

- Learning ML/DL/AI is not equal to learning tools
- The project you work on should guide the tool you use, and not vice-versa
- **Diversifying tools/projects** often makes for a strong profile -> better candidate for industrial jobs

- I used **Tensorflow** for some undergrad projects on relation extraction, activity detection, word embeddings, etc.
- Used **PyTorch** for quick speaker recognition evaluation last summer
- Mostly use **Kaldi** (C++ based speech recognition toolkit) for research
- Extended **scikit-learn**’s clustering package for a project this summer

But there is so much out there!

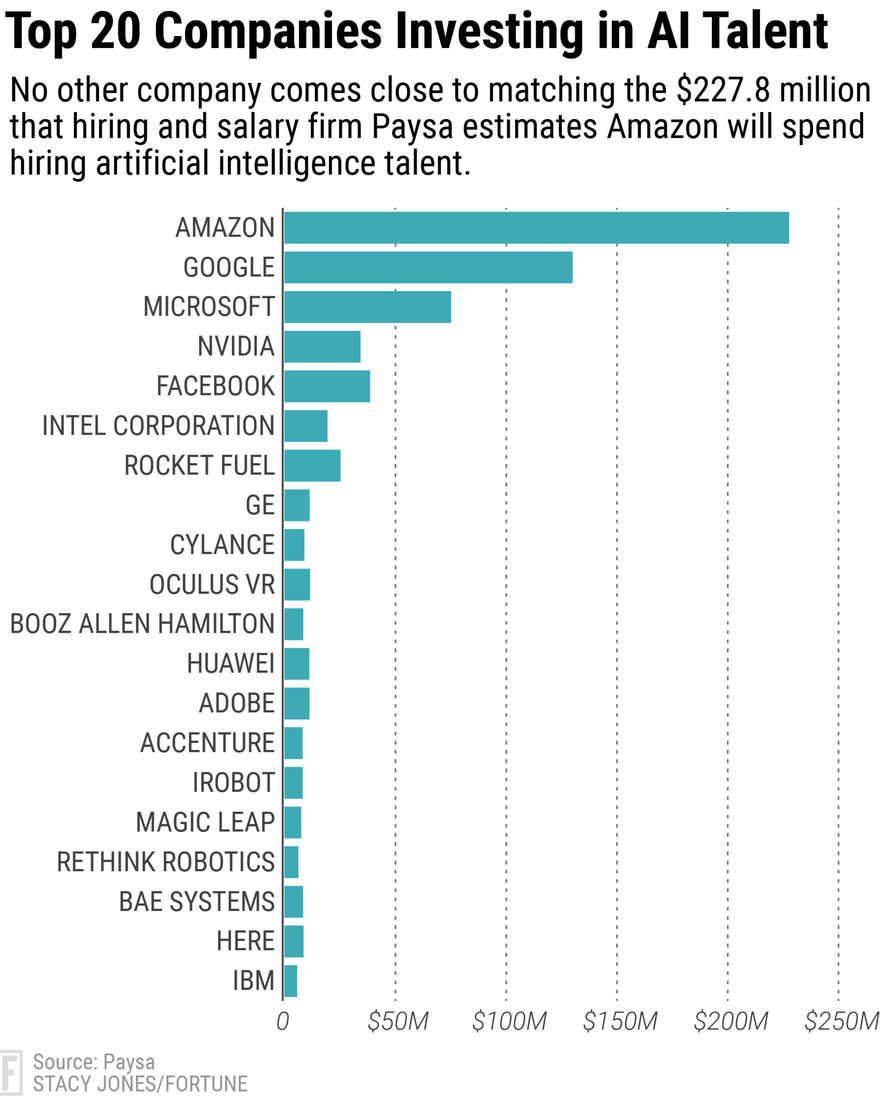
- Learning ML/DL/AI is not equal to learning tools
- The project you work on should guide the tool you use, and not vice-versa
- **Diversifying tools/projects** often makes for a strong profile -> better candidate for industrial jobs

- I used **Tensorflow** for some undergrad projects on relation extraction, activity detection, word embeddings, etc.
- Used **PyTorch** for quick speaker recognition evaluation last summer
- Mostly use **Kaldi** (C++ based speech recognition toolkit) for research
- Extended **scikit-learn**'s clustering package for a project this summer

Libraries are to machine learning what programming languages are to computer science.

Job prospects for NLP/Speech research

- All big companies (FAANG) are investing heavily in AI research, in NLP, speech, and vision.
- Lots of smaller startups aiming at particular applications. Examples:
  - **Duolingo**: translation, language learning
  - **Grammarly**: parsing, syntax, etc.
  - **Huggingface**: all things NLP
  - **Otter.ai, Mobvoi, Notable**: speech recognition
  - **Omilia, Dialogflow**: conversational agents



Job prospects for NLP/Speech research

Profile	Job description	Usually requires
Software developer	Build user-facing features	B.Tech./BE/BS
Research engineer	Implement/deploy models for production	MS/PhD (or BE + experience)
Research scientist	Build new models/ algorithms	PhD (or occasionally MS)

Job prospects for NLP/Speech research

Profile	Job description	Usually requires
Software developer	Build user-facing features	B.Tech./BE/BS
Research engineer	Implement/deploy models for production	MS/PhD (or BE + experience)
Research scientist	Build new models/ algorithms	PhD (or occasionally MS)

- **Grad school** is a great option for specializing in NLP/ speech research
- **Big tech companies** hire MS/PhD students for research positions
- But good research **jobs at startups** also open for undergrads (especially with strong profile)

Thanks!

Contact: r.desh26@gmail.com