

KAIST Summer Session 2018

Module 3. Deep Learning with PyTorch

Leveraging Deep Learning on Your Computer

**KAIST College of Business** 

Jiyong Park

30 August, 2018



#### **Deep Learning Requires Intense Resource**

• Example: Exploration for neural machine translation (Britz et al. 2017)

## Massive Exploration of Neural Machine Translation Architectures

Denny Britz\*, Anna Goldie, Minh-Thang Luong, Quoc Le {dennybritz, agoldie, thangluong, qvl}@google.com
Google Brain

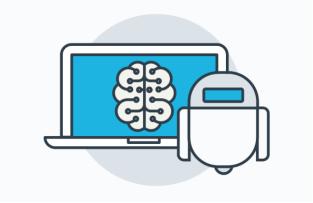
"One major drawback of current architectures is that they are expensive to train, typically requiring days to weeks of GPU time to converge."

We report empirical results and variance numbers for several hundred experimental runs, corresponding to **over 250,000 GPU hours** on the standard WMT English to German translation task... **consisting of 4.5M sentence pairs**"

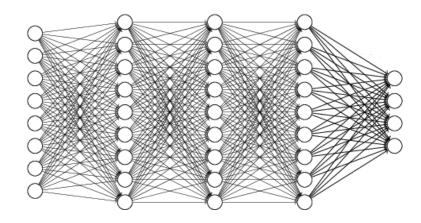


Britz, D., Goldie, A., Luong, T. and Le, Q., 2017. Massive Exploration of Neural Machine Translation Architectures. arXiv preprint arXiv:1703.03906.

## Can We Leverage Deep Learning on Our Computer?



Your Labtop



Deep Learning



Big Data





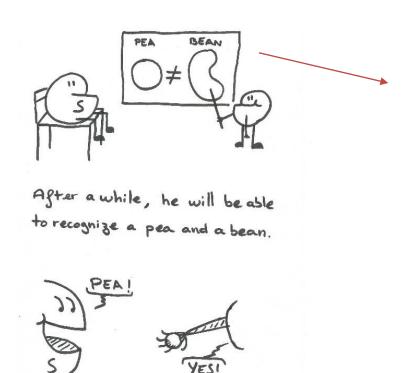
## **Amazon Mechanical Turk**





## **Supervised Learning Needs Labeled Data**

- Most research contexts are lacking labeled data.
  - ➤ How can we generate labeled dataset to train a supervised learning?



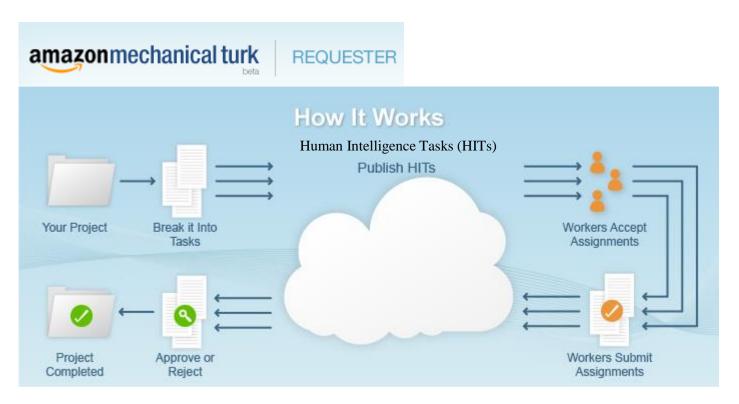
Inputs	Desired outputs
Round shape	Pea
Egg shape	Bean
Round shape	Pea
Round shape	Pea





#### **How to Label Your Data**

 Amazon Mechanical Turk (MTurk) has been very popular in academic research for the goal of data labeling or tagging.



https://requester.mturk.com/





#### **How to Label Your Data**

- Example: Lee et al. (2018) (See Module 1.8 for more examples)
  - Research question

Does the textual advertising content influence consumer engagement on social media?

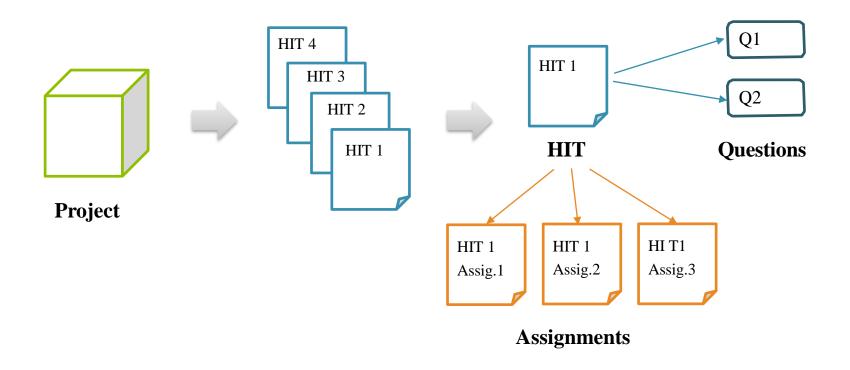
- Economining approach
  - (1) [Labeled data] Human-codingthrough Amazon Mechanical Turk for5,000 advertising contents on Facebook
  - (2) [Supervised learning] Ensemble learning for content tagging to combine results from the multiple classifiers
  - (3) [Prediction] 100,000 text messages

Sample Messages	Content Tags
Welcome to the unveiling of the Pro Staff RF97 that I co-designed with Wilson	BRANDMENTION,
Tennis. Learn more at http://bit.ly/29JXLdA. #FromFederer	PRODMENTION, PRODLOCATION, HTTP
Coach Seve and me. Excited to be back in Brisbane! Happy we got the 1st practice of year out of the way!	SMALLTALK, EMOTION
Hello fans from Colombia! I am very happy to see you at the exo I am playing vs Tsonga on Saturday, December 15th! Buy your tickets starting September 12th on www.tuboleta.com. I hope to see you all there!	EMOTION, SMALLTALK, TARGET, PRODAVAIL, PRODLOCATION, PRODMENTION, HTTP
The Walking Dead Season 1 DVD/Blu-ray is now available, purchase it now!!! http://blogs.amctv.com/the-walking-dead/2011/03/season-1-dvd-blu-ray.php	BRANDMENTION, PRODMENTION, PRODAVAIL, PRODLOCATION, HTTP
Daryl makes a funny. What are some of your favorite #TheWalkingDead quotes? The highest rated quote will be turned into a graphic! #tbt	SMALLTALK, EMOTION, QUESTION, BRANDMENTION, ASKCOMMENT





#### **MTurk Basic Elements**

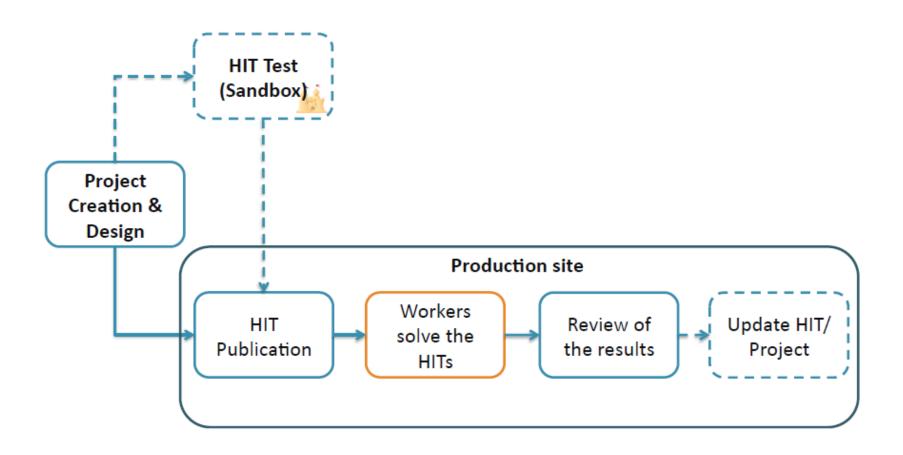


Total cost of the project = No. of HITs × No. of Assignments × (Reward per HIT + Fee)





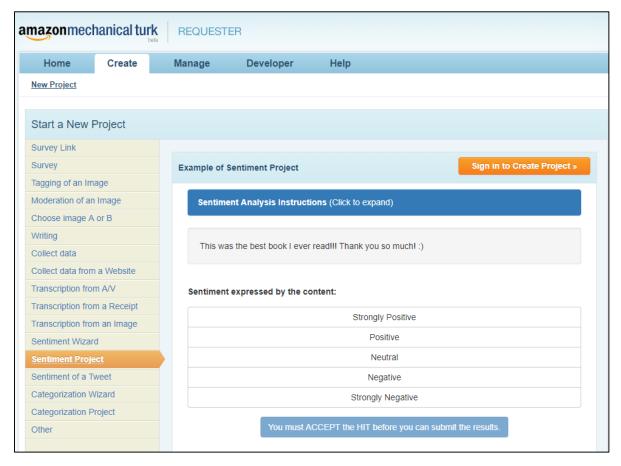
## **MTurk Workflow for Requesters**







• (1) Create a project using predefined templates

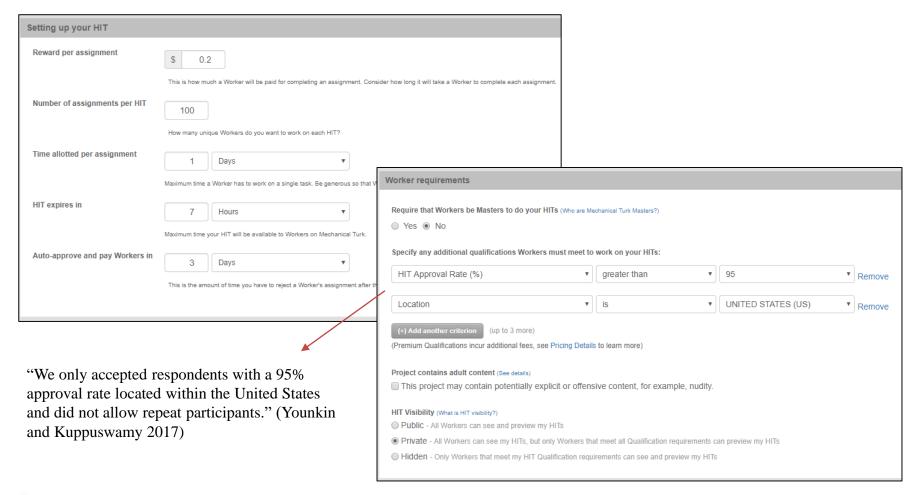


https://requestersandbox.mturk.com/





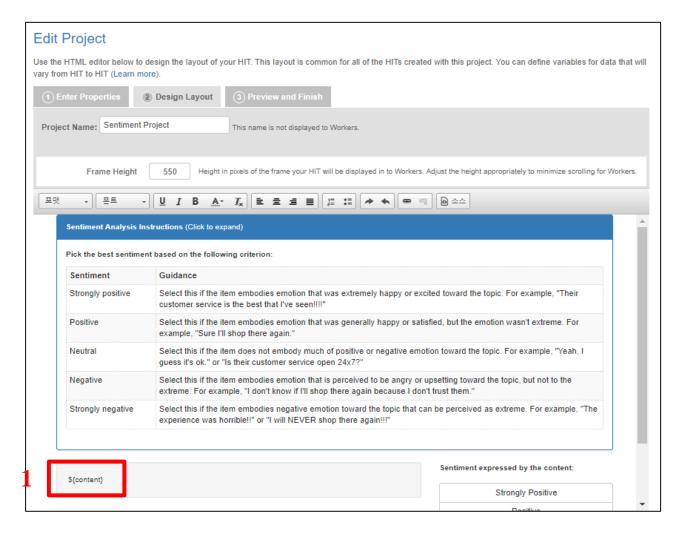
• (2) Set up the HIT properties







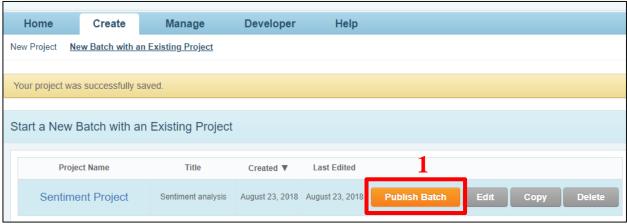
• (3) Design the project

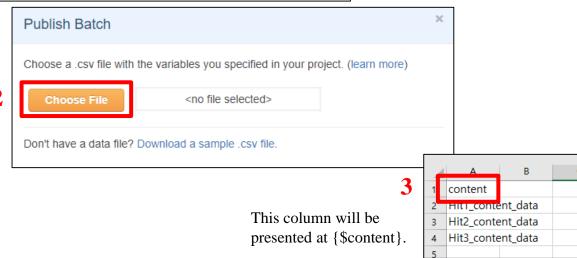






• (4) Publish the HIT









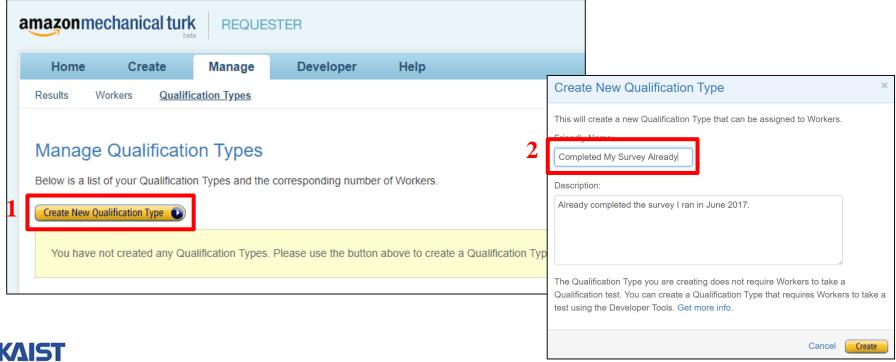
#### • (4) Publish the HIT

Batch Properties					
Title:	Sentiment analysis				
Description:	Analyze the sentiment of the provided content based on the given criteria				
Batch expires in:	7 Hours				
Results are auto-approved and Workers are paid after:	3 Days				
Workers must meet the following Qualifications to work on these HITs:					
	HIT Approval Rate (%) score greater than 95				
	Location score is <u>US</u>				
HITs					
Number of HITs in this batch:		3			
Number of assignments per HIT:	Х	100			
Total number of assignments in this batch:		300			
Cost Summary					
Reward per Assignment:		\$0.20			
	Х	300	(total number of assignments in this batch)		
Estimated Total Reward:		\$60.00	-		
Estimated Fees to Mechanical Turk:	+	\$24.00	(fee details)		
Estimated Cost:		\$84.00	-		
Applied Prepaid HITs Balance:	-	\$84.00			
Remaining Balance Due:		\$0.00	-		



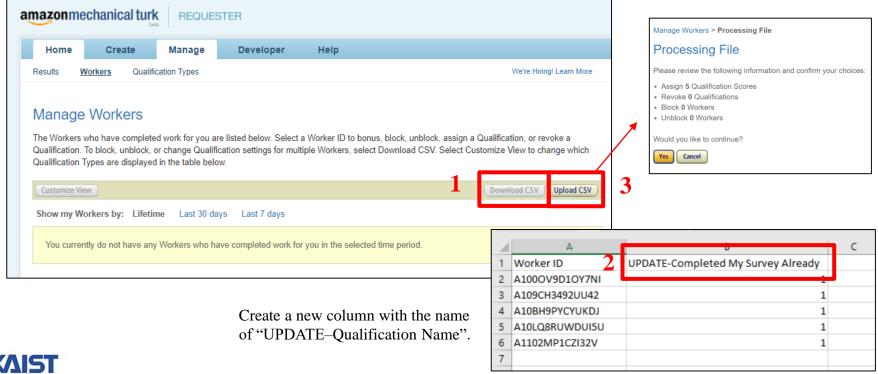


- (5) Useful tips Divide your work into some pieces
  - For pilot tests, it is recommended to publish the HITs with smaller assignment.
  - ➤ (IMPORTANT) The turkers who participated in previous HITs should not be allowed to participate in the subsequent HITs. It is also used for follow-up survey.





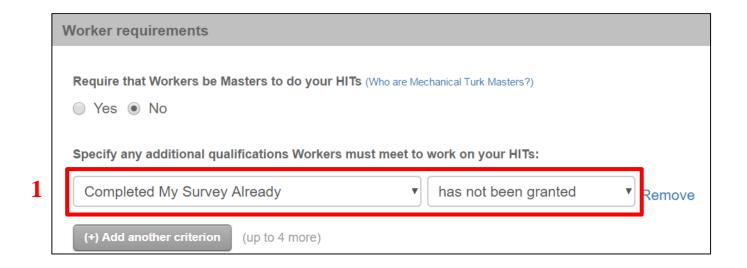
- (5) Useful tips Divide your work into some pieces
  - For pilot tests, it is recommended to publish the HITs with smaller assignment.
  - ➤ (IMPORTANT) The turkers who participated in previous HITs should not be allowed to participate in the subsequent HITs. It is also used for follow-up survey.







- (5) Useful tips Divide your work into some pieces
  - For pilot tests, it is recommended to publish the HITs with smaller assignment.
  - ➤ (IMPORTANT) The turkers who participated in previous HITs should not be allowed to participate in the subsequent HITs. It is also used for follow-up survey.







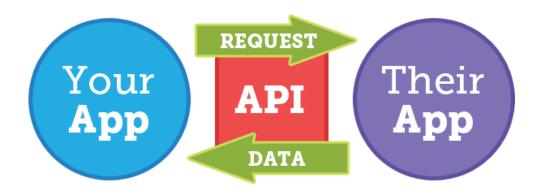
## **How to Use Cloud-based APIs**





#### What is an API?

- API is the acronym for Application Programming Interface, which is a software intermediary that allows two applications to talk to each other.
  - Each time you use an app like Facebook, send an instant message, or check the weather on your phone, you're using an API.
  - ➤ Recently, many IT companies provide cloud-based APIs for deep learning applications (e.g., <u>Google Cloud</u>, <u>Microsoft Azure</u>, <u>IBM Watson</u>)

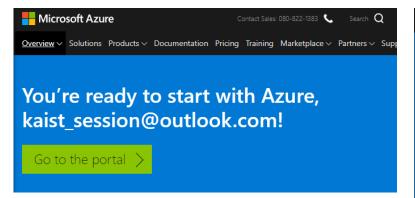


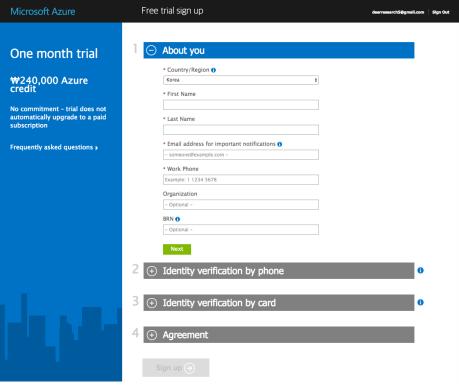




## **Getting an API Key**

• Microsoft Azure (<a href="https://azure.microsoft.com/en-us/free/">https://azure.microsoft.com/en-us/free/</a>)





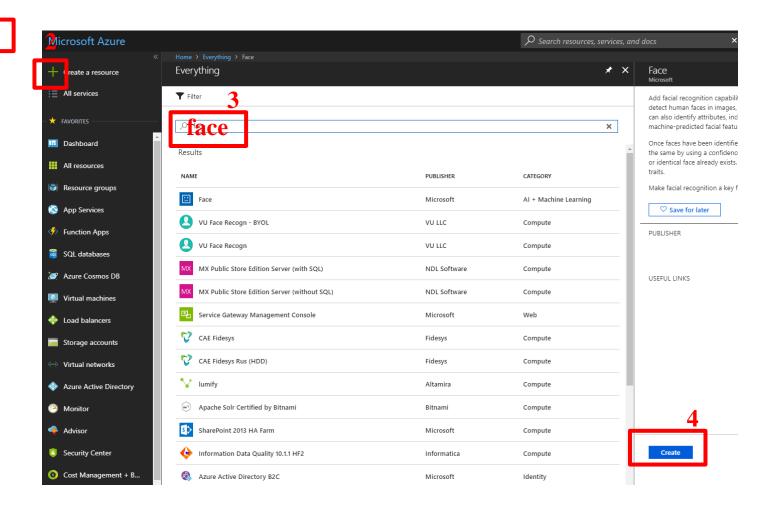




## **Getting an API Key**

1

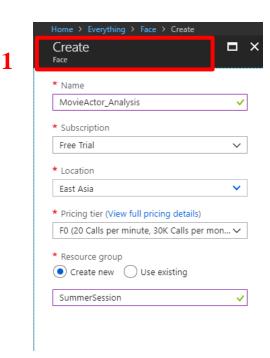
portal.azure.com

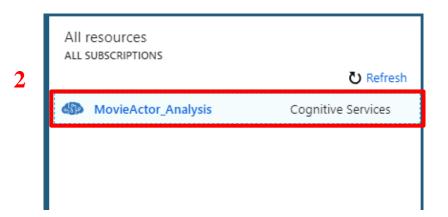


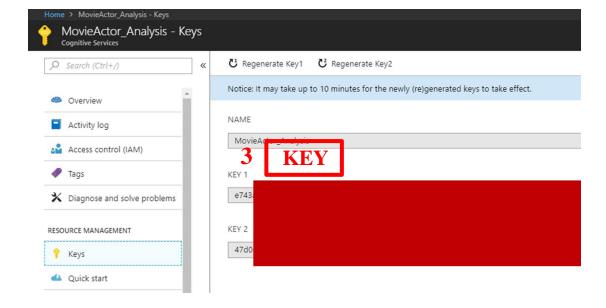




## **Getting an API Key**





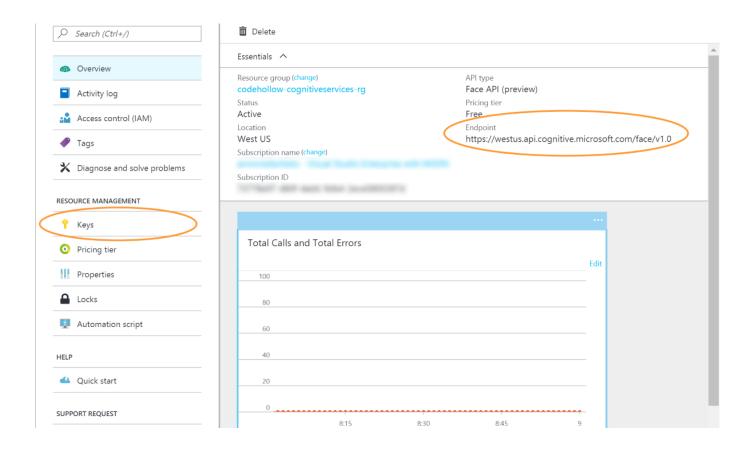






## Getting an Endpoint URL

The endpoint URL you request to API may be different based on location.









#### **How to Use API**

M3.8 MS Face API\_Face Detection.ipynb





## **Running Deep Learning on the Cloud**



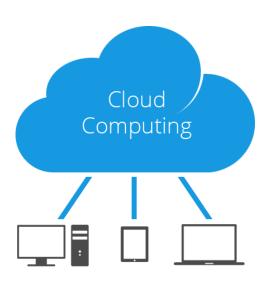


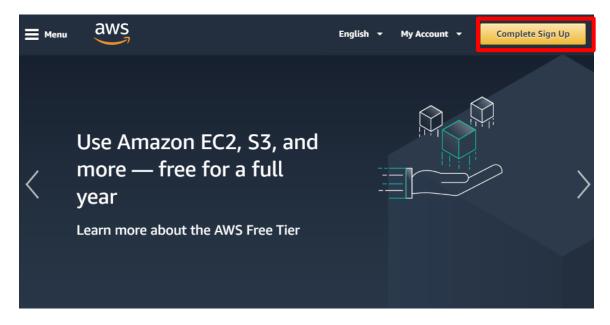
## **Cloud Computing**

• We can run deep learning applications on a virtual machine located in cloud computing which provides shared pools of configurable system resources (e.g., Amazon Web Service).

https://aws.amazon.com/

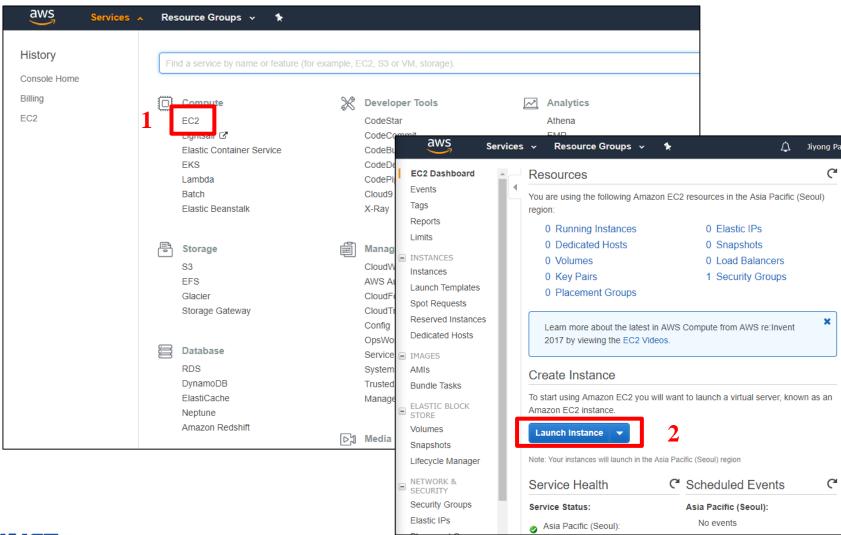
2





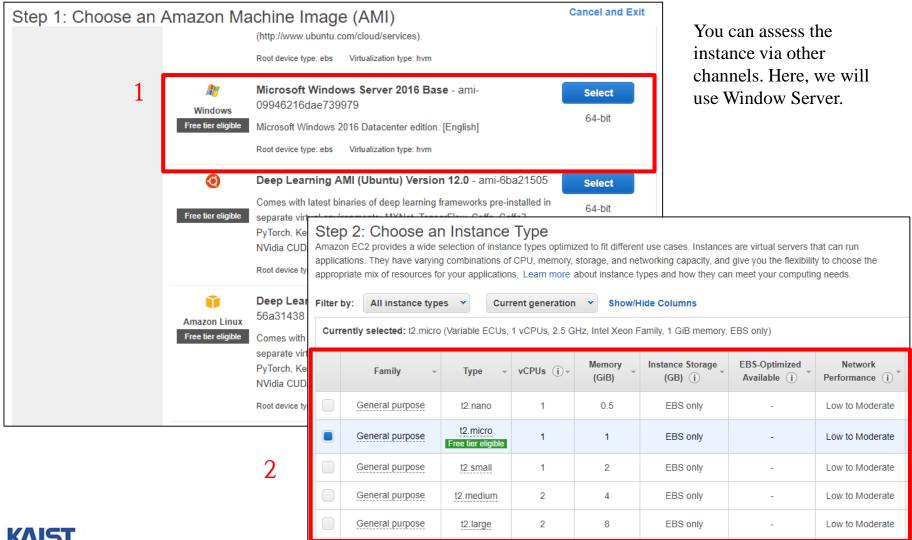














#### Amazon EC2 Instance Types

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your applications. Each instance type includes one or more instance sizes, allowing you to scale your resources to the requirements of your target workload.

Get Started with AWS for Free

Create a Free Account

AWS Free Tier includes 750 hours of Linux and Windows t2.micro instances each month for one year. To stay within the Free Tier, use only EC2 Micro instances.

View AWS Free Tier Details »

General Purpose Compute Optimized Memory Optimized

Accelerated Computing Storage Optimized

Choose the specifications of virtual machine for your purpose.

#### P3

P3 instances are the latest generation of general purpose GPU instances.

#### Features:

- Up to 8 NVIDIA Tesla V100 GPUs, each pairing 5,120 CUDA Cores and 640 Tensor Cores
- High frequency Intel Xeon E5-2686 v4 (Broadwell) processors
- · Supports NVLink for peer-to-peer GPU communication
- Provide Enhanced Networking using Elastic Network Adapter with up to 25 Gbps of aggregate network bandwidth within a Placement Group

Model	GPUs	vCPU	Mem (GiB)	GPU Mem (GiB)	GPU P2P
p3.2xlarge	1	8	61	16	-
p3.8xlarge	4	32	244	64	NVLink
p3.16xlarge	8	64	488	128	NVLink

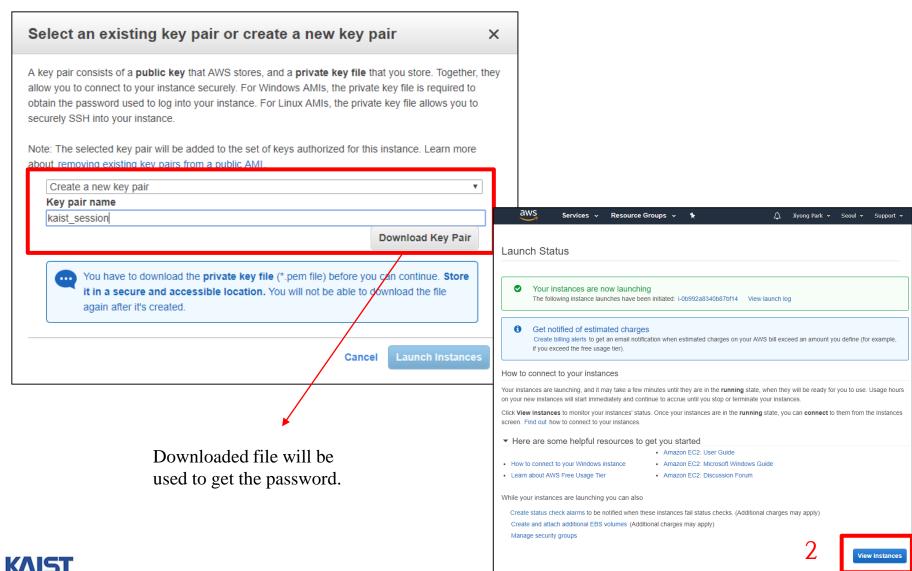
#### **Use Cases**

Machine/Deep learning, high performance computing, computational fluid dynamics, computational finance, seismic analysis, speech recognition, autonomous vehicles, drug discovery.

https://aws.amazon.com/ec2/instance-types





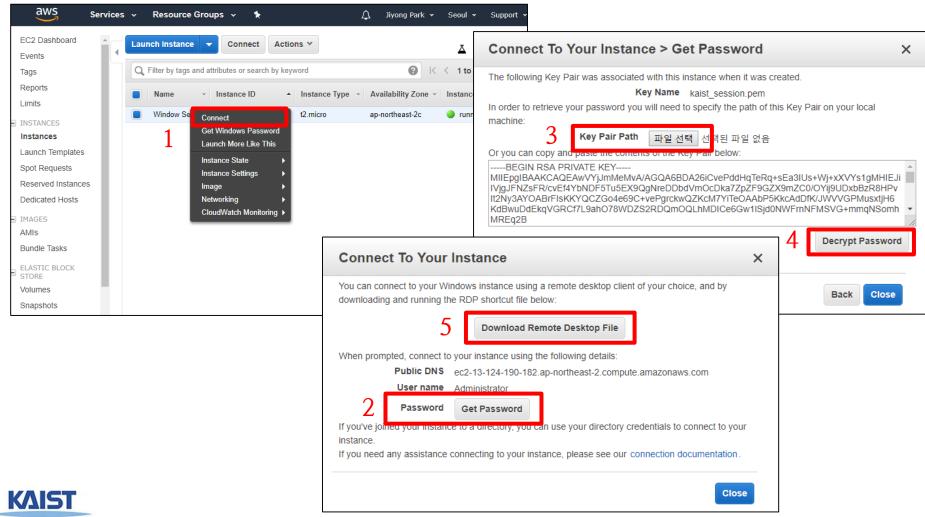




#### **Access Instance (Virtual Machine)**

For Windows users

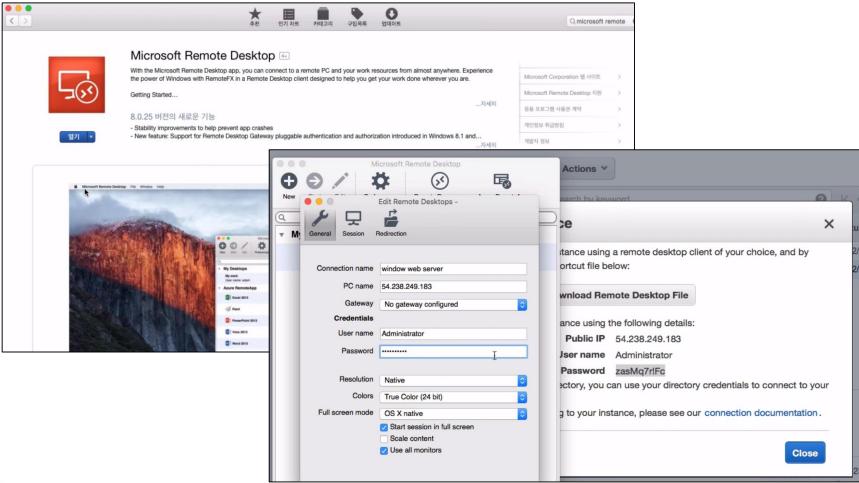
**COLLEGE OF BUSINESS** 





### **Access Instance (Virtual Machine)**

For macOS users, you may use "Microsoft Remote Desktop"







### **Access Instance (Virtual Machine)**

- You can assess a new window running on the cloud.
  - Install Python and keep training your model





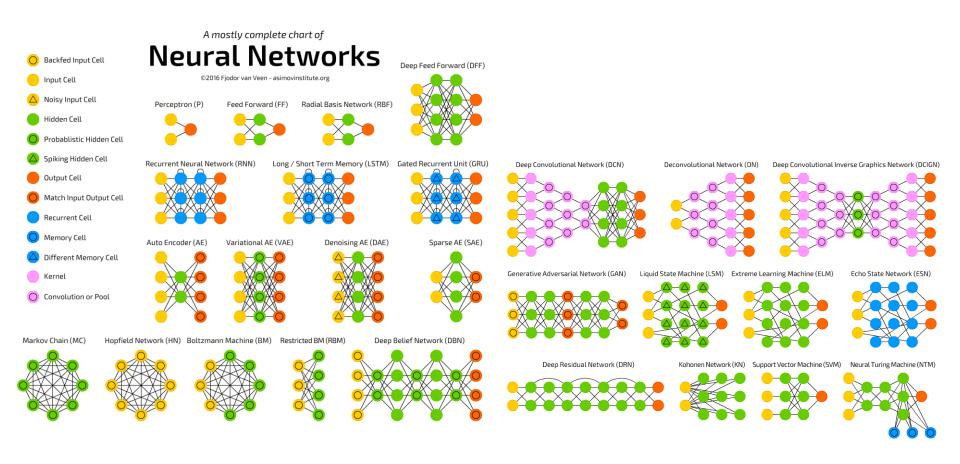


## **Conclusion**





## **Enjoy Your Deep Learning!**





https://becominghuman.ai/cheat-sheets-for-ai-neural-networks-machine-learning-deep-learning-big-data-678c51b4b463



# End of Document

