



Google AI ML Winter Camp
谷歌 AI 机器学习应用冬令营

Project Announcement

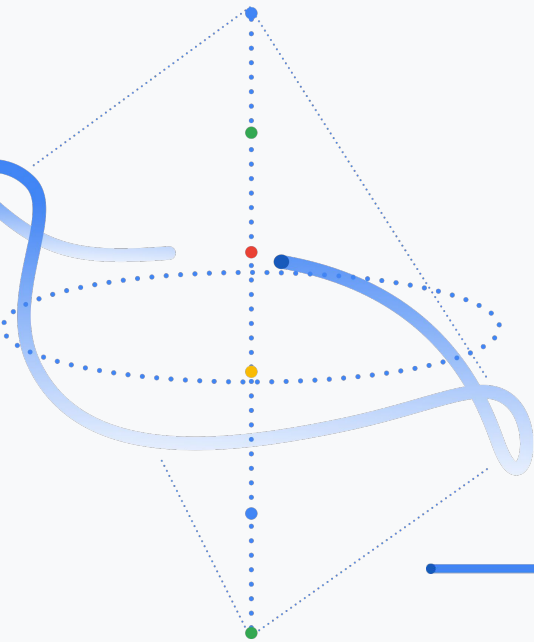
Google AI ML Winter Camp, PEK
Jan 21 -- 25, 2019

ML Projects in Action

Practice real ML application

ML requires practice with real problems.

- Form: Group project (≤ 3 ppl)
- Complete an innovative ML application.
 - Pick a project from candidate list, and formulate an ML problem (CV, NLP, etc.).
 - Design an ML algorithm.
 - Train models and iterate to tune parameters.
 - Evaluation and application demo.
- Github repo (you own your project, but open for peer review)
- Friday: Presentation and poster.

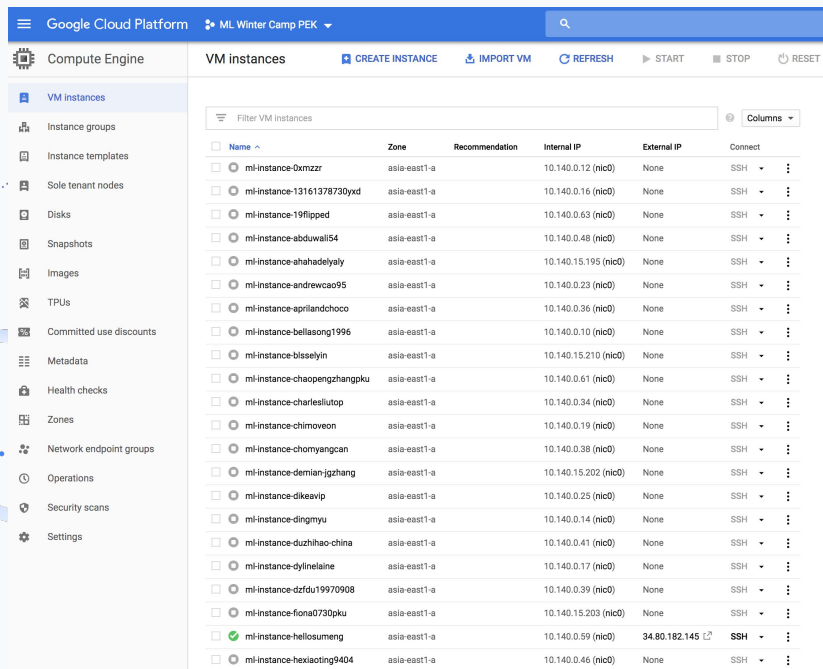


Computational Resources

Each one of you is provided with 1 GCP Linux machine.

<https://console.cloud.google.com/home>

Project ID: **ml-winter-camp-pek**



The screenshot shows the Google Cloud Platform console interface. The left sidebar contains navigation links for Compute Engine, VM instances, Instance groups, Instance templates, Sole tenant nodes, Disks, Snapshots, Images, TPUs, Committed use discounts, Metadata, Health checks, Zones, Network endpoint groups, Operations, Security scans, and Settings. The main panel displays the 'VM instances' page for the project 'ML Winter Camp PEK'. It includes a search bar, a 'Filter VM instances' dropdown, and a table of VM instances. The table has columns for Name, Zone, Recommendation, Internal IP, External IP, and Connect. The 'mi-instance-hellosumeng' instance is highlighted with a green checkmark in the 'Connect' column.

Name	Zone	Recommendation	Internal IP	External IP	Connect
mi-instance-0xmzr	asia-east1-a		10.140.0.12 (nic0)	None	SSH
mi-instance-13161378730yxd	asia-east1-a		10.140.0.16 (nic0)	None	SSH
mi-instance-19flipped	asia-east1-a		10.140.0.63 (nic0)	None	SSH
mi-instance-abduwall54	asia-east1-a		10.140.0.48 (nic0)	None	SSH
mi-instance-ahahadelaly	asia-east1-a		10.140.15.195 (nic0)	None	SSH
mi-instance-andrewcao95	asia-east1-a		10.140.0.23 (nic0)	None	SSH
mi-instance-aprilandchoco	asia-east1-a		10.140.0.36 (nic0)	None	SSH
mi-instance-bellason1996	asia-east1-a		10.140.0.10 (nic0)	None	SSH
mi-instance-blisselyn	asia-east1-a		10.140.15.210 (nic0)	None	SSH
mi-instance-chaopengzhangku	asia-east1-a		10.140.0.61 (nic0)	None	SSH
mi-instance-charlesluto	asia-east1-a		10.140.0.34 (nic0)	None	SSH
mi-instance-chimoveon	asia-east1-a		10.140.0.19 (nic0)	None	SSH
mi-instance-chomyangcan	asia-east1-a		10.140.0.38 (nic0)	None	SSH
mi-instance-demianjgzhang	asia-east1-a		10.140.15.202 (nic0)	None	SSH
mi-instance-dikeavip	asia-east1-a		10.140.0.25 (nic0)	None	SSH
mi-instance-dingmyu	asia-east1-a		10.140.0.14 (nic0)	None	SSH
mi-instance-duzhihao-china	asia-east1-a		10.140.0.41 (nic0)	None	SSH
mi-instance-dylinelaine	asia-east1-a		10.140.0.17 (nic0)	None	SSH
mi-instance-dzdfdu19970908	asia-east1-a		10.140.0.39 (nic0)	None	SSH
mi-instance-fiona0730pku	asia-east1-a		10.140.15.203 (nic0)	None	SSH
mi-instance-hellosumeng	asia-east1-a		10.140.0.59 (nic0)	34.80.182.145	SSH
mi-instance-hexiaoting9404	asia-east1-a		10.140.0.46 (nic0)	None	SSH

```
Linux admin-instance 4.9.0-8-amd64 #1 SMP Debian 4.9.110-3+deb9u6 (2018-10-08) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

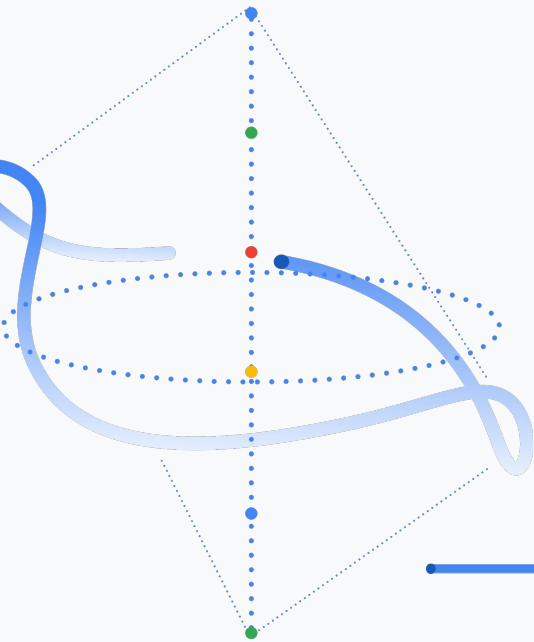
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sun Jan 13 14:13:54 2019 from 74.125.41.97
tianlin@admin-instance:~$
```

Computational Resources

Each one is provided with 1 GCP Linux machine:

- 16 CPUs
- 60 GB memory
- 1 NVIDIA Tesla P100 GPU
- 500 GB SSD persistent disk

During Winter Camp. (Mon. -- Fri.; Backup Sat.)



Project List

Food 101 (CV)

https://www.vision.ee.ethz.ch/datasets_extra/food-101/

Data

101 food categories. 750 training image for each category and 250 test images for each category.

(Downloaded: <gs://ml-winter-camp-pek/food-101>. [browse](#))

Potential Tasks

Food recognition, similarity search.



Project List

Face Attribute Transfer (CV)

<http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>

Data

10,177 identities, 202,599 face images, and 5 landmark locations, 40 binary attributes annotations
(Downloaded: <gs://ml-winter-camp-pek/CelebA>. [browse](#))

Potential Tasks

Image Transfer, Face Recognition

Eyeglasses



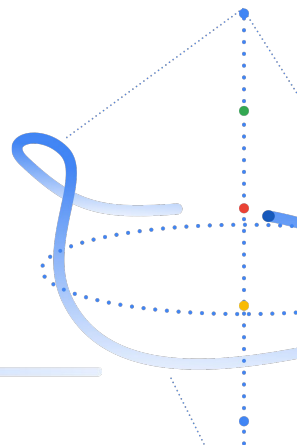
Wearing Hat



Bangs



Wavy Hair



Project List

Cartoon Face (CV)

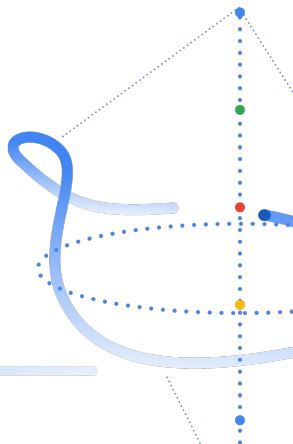
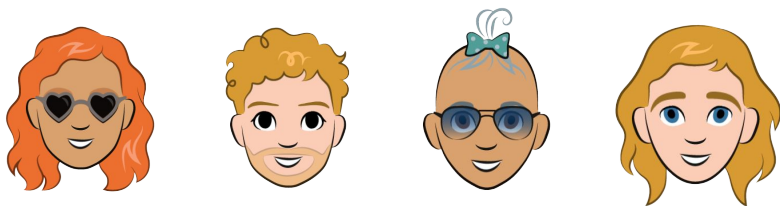
<https://google.github.io/cartoonset/download.html>

Data

10k/100k images. 16 components that vary in 10 artwork attributes (eye, chin, face, hair, etc.), 4 color attributes, and 4 proportion attributes.

Potential Tasks

Face recognition, classification, transfer



Project List

iMaterialist Challenge (Fashion) (CV)

<https://www.kaggle.com/c/imaterialist-challenge-fashion-2018/data>

Data

1+ million products with 228 fashion attribute classes: color, style, pattern, etc. *(with additional label map)*

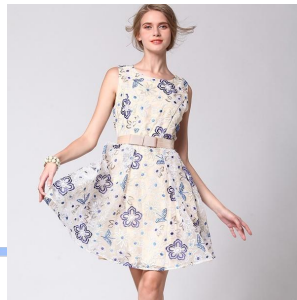
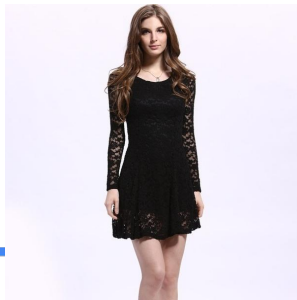
Potential Tasks

(Downloaded: <gs://ml-winter-camp-pek/fashion>. [browse](#))

Image classification, similarity search

Doc

See: Project Data - Fashion - Label Map *(attached doc)*



Project List

Speech Synthesis (Speech)

<https://keithito.com/LJ-Speech-Dataset/>

Data

Public domain speech dataset. 13,100 short audio clips of a single speaker reading passages from 7 non-fiction books. 1~10 seconds each, 24 hours in total.

Potential Tasks

(Downloaded: <gs://ml-winter-camp-pek/SpeechSynthesis>. [browse](#))

Text To Speech (Char2Feats, WaveRNN)

Instructions

Use <https://github.com/tensorflow/lingvo>

See: **Project Instructions - Build a Text-To-Speech System** (attached doc)

Sample Audio:

Many animals of even complex structure which live parasitically within others are wholly devoid of an alimentary cavity.

Project List

Humpback Whale Identification (CV)

<https://www.kaggle.com/c/humpback-whale-identification>

Data

25,000+ images with 3,000+ whale IDs gathered from research institutions and public contributors. Real world problem for animal protection.

Potential Tasks

Image classification, similarity search



Project List

Google Play Store Apps (NLP, Structure Data)



<https://www.kaggle.com/lava18/google-play-store-apps/data>


Data

Play store data with category, rating, comments, etc. 10K apps, 60K reviews.

Potential Tasks

Structure data or text classification, regression, multi-task learning

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres
Application name	Category the app belongs to	Overall user rating of the app (as when scraped)	Number of user reviews for the app (as when scraped)	Size of the app (as when scraped)	Number of user downloads/installs for the app (as when scraped)	Paid or Free	Price of the app (as when scraped)	Age group the app is targeted at - Children / Mature 21+ / Adult	An app can belong to multiple genres (apart from its main category). For eg. a musical game will belong to
9660 unique values	FAMILY 18% GAME 11% Other (32) 71%			Varies with device 16% 11M 2% Other (460) 83%	1,000,000+ 15% 10,000,000+ 12% Other (20) 74%	Free 93% Paid 7% Other (2) 0%	0 93% \$0.99 1% Other (91) 6%	Everyone 80% Teen 11% Other (5) 8%	Tools Entertainment Other (118)
1 Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design
2 Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pre-Play
3 U Launcher Lite - FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87518	8.7M	5,000,000+	Free	0	Everyone	Art & Design
4 Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design

App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subject
Name of app	User review (Preprocessed and translated to English)	Positive/Negative/Neutral (Preprocessed)	Sentiment polarity score	Sentiment subjectivity score
1074 unique values	nan 42% Good 0% Other (27994) 58%	nan 42% Positive 37% Other (2) 21%		nan 42% 0.0 7% Other (4693) 51%
10 Best Foods for You	I like eat delicious food. That's I'm cooking food myself, case "10 Best Foods" helps lot, also "Best Before (Shelf Life)".	Positive	1.0	0.5333333333333333
10 Best Foods for You	This help eating healthy exercise regular basis	Positive	0.25	0.28846153846153844

Project List

Reddit Comments (NLP)

<https://www.kaggle.com/danofer/sarcasm/data>

Data

1.3 million comments with 5 scores, up & down numbers.

Potential Tasks

Text classification, regression, multi-tasks

	# label	A comment	A author	A subreddit	# score	# ups	# downs	📅 date	📅 created...	A parent_...
1	0	NC and NH.	Trumpbart	politics	2	-1	-1	2016-10	2016-10-16 23:55:23	Yeah, I get that argument. At this point, I'd prefer is she lived in NC as well.
2	0	You do know west teams play against west teams more than east teams right?	Shbshb906	nba	-4	-1	-1	2016-11	2016-11-01 00:24:10	The blazers and Mavericks (The wests 5 and 6 seed) did not even carry a good

Project List

WikiJoin (NLP)

gs://ml-winter-camp-pek/SpeechSynthesis. [browse](#)

Data

114,975 paired articles in simple/full versions. Article has structure information: headings and contents in sections.

Potential Tasks

Text classification, text summarization

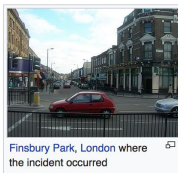
Doc

See: Project Data - WikiJoin (*attached doc*)

On 19 June 2017, at 00:15 BST, police officers responded to a vehicle running over several civilians with a reported "numerous casualties" in [Finsbury Park, London](#).^[1] The attack occurred near [Finsbury Park Mosque](#) where the van ran over several worshippers.^[2]

The car reportedly crashed close to the Finsbury Park Mosque, with many of the injured seen wearing Muslim attire. Witnesses have said this is because worshippers were leaving the mosque at the time of the incident.^[3] One person has been arrested in relation to the attack. The individual, Darren Osborne, has been confirmed to be the attacker. It is being reported that 10 people have been injured in the attack and one person has been confirmed to have died.^[4]

A witness reported to [BuzzFeed News](#) that the driver screamed that he wanted to "kill all Muslims".^[5]



References

- ¹ "Finsbury Park: 'Several hurt' as vehicle hits pedestrians" [@](#). BBC News. 19 June 2017. Retrieved 18 June 2017.
- ² "Van hits worshippers outside Finsbury Park Mosque in London" [@](#). Independent. 19 June 2017. Retrieved 18 June 2017.
- ³ "Finsbury Park: 'Several hurt' as vehicle hits pedestrians" [@](#). 19 June 2017. Retrieved 19 June 2017 – via [www.bbc.com](#).
- ⁴ "Van hits pedestrians in London; 'number of casualties' reported" [@](#). Tucson News Now. 18 June 2017. Retrieved 18 June 2017.
- ⁵ "A Vehicle Has Struck Worshippers Outside A London Mosque, Causing Injuries" [@](#). BuzzFeed News. June 18, 2017. Archived from the original [@](#) on June 19, 2017. "He was screaming before that — I'm going to kill all Muslims," Saleh Aloudi said. "He was throwing punches all over."

2017 Finsbury Park attack

From Wikipedia, the free encyclopedia

Coordinates: 51°56′30″N 0°10′83″W﻿ / ﻿

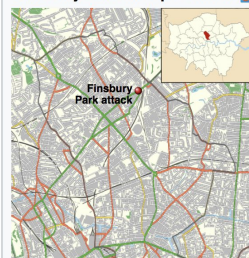
The **Finsbury Park attack** was a vehicle-ramming attack in [Finsbury Park](#), London, England, on 19 June 2017. A van was [driven into pedestrians](#) in Finsbury Park, London, by Darren Osborne, causing one death and injuring at least nine people. This occurred near the [Muslim Welfare House](#), 100 yards (90 m) from [Finsbury Park Mosque](#). A man who had earlier collapsed and was receiving [first aid](#) died at the scene. The attack followed a series of other [car-ramming attacks](#) that happened during the year of 2017.

The incident was investigated by [counter-terrorism](#) police as a terrorist attack.^{[5][6][7]} On 23 June, Darren Osborne of [Cardiff](#) was charged with terrorism-related murder and attempted murder.^{[8][9]} In early February 2018 at [Woolwich Crown Court](#), he was found guilty on both counts^[2] and was sentenced to life imprisonment.^[10]

Contents [[hide](#)]

1 [Background](#)

Finsbury Park Mosque attack



Project List

Chinese Daily New Text (NLP)









<https://www.kaggle.com/noxmoon/chinese-official-daily-news-since-2016>

Data

Chinese news data, Xinwen Lianbo. 20,738 piece of news since 1 January 1978.

Potential Tasks

Text summarization/abstraction, transfer learning

 date		 tag		 headline		 content	
date of the news		“详细全文” some relatively long news, “国内” domestic short news, “国际” international short news. These tags may not		summary for each piece of news		news text	

Project List

MBTI personality type with posts (NLP)

<https://www.kaggle.com/datasnaek/mbti-type>

Data

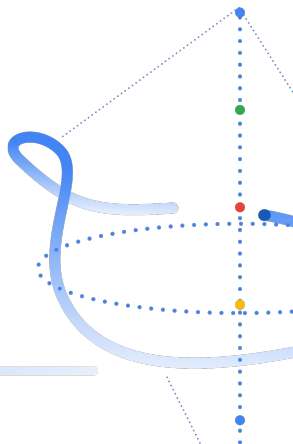
The Myers Briggs Type Indicator (or MBTI for short). 16 personality types across 4 axis. Text Classification 8675 examples

Potential Tasks

Text classification, sentimental analysis

I (Introversion) N (Intuition) F (Feeling) P(Perceiving)

... Hey there stranger :P I remember you from when I first joined PerC several years ago. I always felt you were cheerful, intelligent, and wise beyond your years. Look at the amount of thanks you...



Project List

Relatively Easy Kaggle Tasks

Data

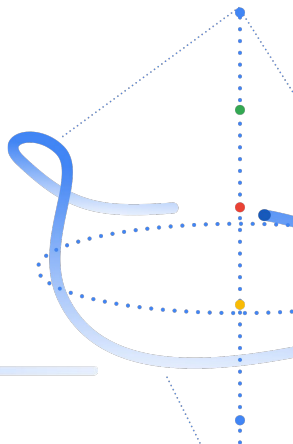
Titanic: <https://www.kaggle.com/c/titanic>

House Prices: <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

Predict Future Sales: <https://www.kaggle.com/c/competitive-data-science-predict-future-sales>

Potential Tasks

Structural data prediction, Multi-tasks.



Open Tool with Topic

China Painting Transfer Using GAN (CV)

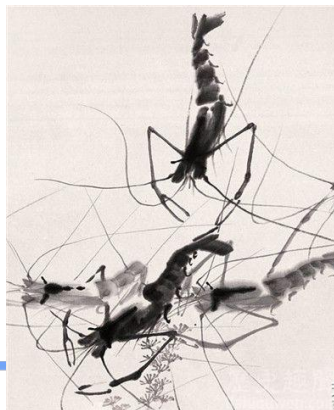
<https://junyanz.github.io/CycleGAN/>

Data

Chinese traditional painting to real photos via CycleGan. For data, **you need to find Chinese paintings** and real photos collection. Original CycleGAN data for coding and debugging.

Tasks






Image Transfer



Open Datasets

Google AI competitions on Kaggle (CV)

<https://www.kaggle.com/googleai/competitions>





5 competitions		
	Quick, Draw! Doodle Recognition Challenge How accurately can you identify a doodle? <i>Featured</i> · a month ago	\$25,000 1,316 teams
	Inclusive Images Challenge Stress test image classifiers across new geographic distributions <i>Research</i> · 2 months ago	\$25,000 468 teams
	Google AI Open Images - Object Detection Track Detect objects in varied and complex images. <i>Featured</i> · 4 months ago	\$30,000 454 teams
	Google AI Open Images - Visual Relationship Track Detect pairs of objects in particular relationships. <i>Featured</i> · 4 months ago	\$20,000 232 teams
	The 2nd YouTube-8M Video Understanding Challenge Can you create a constrained-size model to predict video labels? <i>Featured</i> · 5 months ago	\$25,000 312 teams

Open Datasets

Google BigQuery

<https://www.kaggle.com/bigquery/datasets>

27 datasets

5		USPTO OCE Patent Claims Research Data US patents granted claims and published applications (BigQuery) Google BigQuery (Continuous updates)	research law bigquery	BigQuery 119.3 GB CC4	</> 2 0 967
22		World Development Indicators (WDI) Data World Bank collection of global development indicators (BigQuery) Google BigQuery (Continuous updates)	bigquery economics world finance	BigQuery 2.2 GB CC3	</> 2 0 7k
2		USPTO Patent Trial and Appeal Board (PTAB) Data Trials conducted by PTAB for issues of patentability (BigQuery) Google BigQuery (Continuous updates)	law bigquery	BigQuery 137 MB CC4	</> 2 0 780
2		USPTO Patent Examiner Data System (PEDS) Data Data from the examination process of USPTO patent applications (BigQuery) Google BigQuery (Continuous updates)	bigquery law research	BigQuery 40.2 GB CC4	</> 2 0 877
10		USPTO Patent Examination Research Data (PatEx) Millions of publicly viewable patent applications filed with USPTO (BigQuery) Google BigQuery (Continuous updates)	research law bigquery	BigQuery 30.2 GB CC4	</> 2 0 897
6		The Office Action Research Dataset for Patents Actions taken by patent examiners to patent applicants (BigQuery) Google BigQuery (Continuous updates)	bigquery research law	BigQuery 3.9 GB CC4	</> 2 0 1k

Important Tips:

- Innovative applications are beyond a single model. Find a good problem, create a demo and prepare for presentation.
- End2End asap, and then polish details. Appreciate original ML model. Using API is OK but not major innovation.
- (If based on existing work) we mainly evaluate your work and innovation during Winter Camp. Your reference should be clear.

