

# STUDY OF ENGAGEMENT IN ONLINE LEARNING

Presented by Pony-Unicorn team - 12/10/2020  
Data Science W241  
Experiments and Causal Inference  
UC Berkeley MIDS

# INTRODUCTION

- ✗ Students engagement is important, but hard to measure:
  - ✗ Honest and unbiased evaluation
  - ✗ Enough data to get statistical power
- ✗ Demand for online classes is growing rapidly:
  - ✗ Different challenges for the instructor, compared to in-person
  - ✗ *But also more opportunities?*



# OUR ASPIRATION

- ✗ Engagement measurements
  - ✗ Non-intrusive
  - ✗ Large-scale
  - ✗ Objective
- ✗ Platform for running experiments
  - ✗ Automated
  - ✗ Quality data collection
  - ✗ Statistical analysis





# RESEARCH QUESTION

What is the impact of topical discussions via private chat on student engagement in online classes?

# 1.

# METHODOLOGY



# HYPOTHESIS AND APPROACH

*Interference in online classes has NO impact on student engagements.*

## Key Concepts:

- **Interference:** private IM on the lecture topic
- **Online classes:** zoom platform
- **Students:** volunteers from UC Berkeley MIDS program
- **Engagements:** speech, chat messages and observable face & body signals
- **Measurement tools:** Computer vision & NLP AI



# RESEARCH ETHICS

## w241: Causal Inference

Hey everyone!

Our final group project involves experimental evaluation of engagement in remote education. We are aspiring to do analysis of the zoom video recordings of our class sessions using computer vision and natural language processing AI (open source and / or proprietary code) for our experiment. We want to ask for your permission and consent to avoid any ethical violations. If you don't mind being a part of our experiment, please fill out the form below! We plan on using video recordings from one or more upcoming sessions of 241 class.

All collected data will be anonymized and aggregated when reported, so the experiment will have no bearing on performance in class. We will be using it solely for non-commercial research purpose of delivery and receipt of educational content.

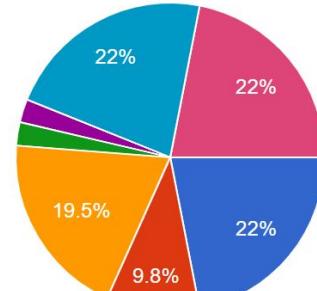
- Bryan, Dana, Casey, Alexandra, Vasanth

\* Required

Which section are you in?

41 responses

- Micah Section 1
- Scott Section 2
- Scott Section 3
- Alex Section 4
- Alex Section 5
- Section 4
- Section 5



We distributed experiment consent forms to students via Google Forms and ran controlled experiments on participants.

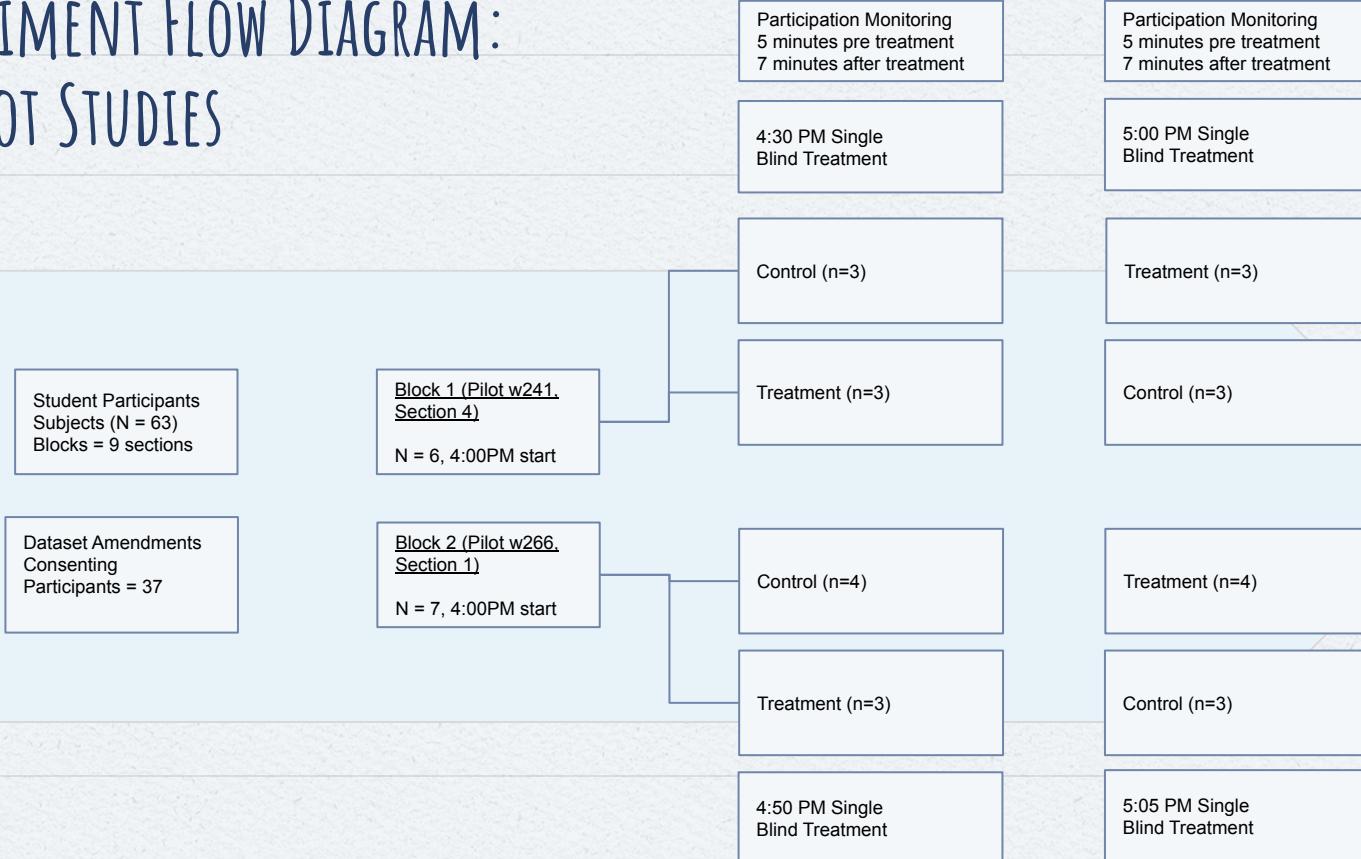
[https://docs.google.com/forms/d/e/1FAIpQLSeME4ouAkWzF7Y4a77WU5TTPpXPRw9lc5\\_HtkRYMpSjBNB9Bw/viewform](https://docs.google.com/forms/d/e/1FAIpQLSeME4ouAkWzF7Y4a77WU5TTPpXPRw9lc5_HtkRYMpSjBNB9Bw/viewform)

# EXPERIMENT DESIGN (PRETEST - POSTTEST CONTROL GROUP DESIGN)

One Session	Subjects (randomly distributed between subgroups)	Treatment (Single-blinded)
Study 1 (randomly picked time)	Subgroup 1 Subgroup 2	R O X O R O -- O
Study 2 (randomly picked time)	Subgroup 1 Subgroup 2	R O -- O R O X O



# EXPERIMENT FLOW DIAGRAM: 2 PILOT STUDIES



DATA  
COLLECTION

1

# TREATMENT AND RANDOMIZATION

\* Experimenter: Dana Kaban

	date	subject	question	experimenter
14	2020-10-29 19:00:00	[REDACTED]	what are we discussing? I missed it.	Dana Kaban

\* Experimenter: Casey King

	date	subject	question	experimenter
3	2020-10-29 18:40:00	[REDACTED]	what are we discussing? I missed it.	Casey King
8	2020-10-29 19:00:00	[REDACTED]	what are we discussing? I missed it.	Casey King

\* Experimenter: Vasanth Ramani

	date	subject	question	experimenter
1	2020-10-29 18:40:00	[REDACTED]	what are we discussing? I missed it.	Vasanth Ramani
2	2020-10-29 18:40:00	[REDACTED]	what are we discussing? I missed it.	Vasanth Ramani
5	2020-10-29 18:40:00	[REDACTED]	what are we discussing? I missed it.	Vasanth Ramani
12	2020-10-29 19:00:00	[REDACTED]	what are we discussing? I missed it.	Vasanth Ramani
15	2020-10-29 19:00:00	[REDACTED]	what are we discussing? I missed it.	Vasanth Ramani

\* Experimenter: Alexandra Savelieva

date subject question experimenter

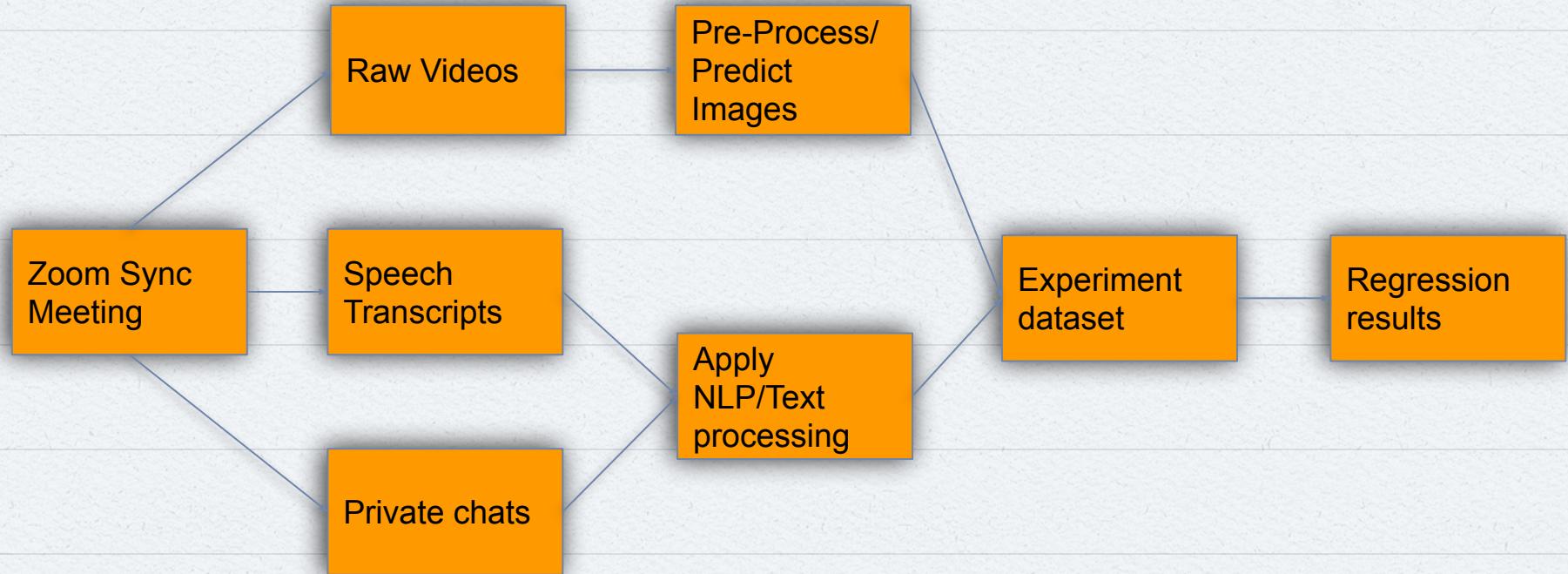
2

18:40:01	From	Vasanth Ramani	to SUBJECT(Privately) :	what are we discussing? I missed it.
18:40:10	From	Vasanth Ramani	to SUBJECT(Privately) :	what are we discussing? I missed it.
18:40:26	From	SUBJECT	to Vasanth Ramani(Privately) :	#5 in ps4
18:40:32	From	Vasanth Ramani	to SUBJECT(Privately) :	what are we discussing? I missed it.
18:40:33	From	SUBJECT	: SUBJECT is saying PS4 problem 5 is crashing on his side	
18:40:58	From	SUBJECT	: sorry, meant to reply privately to Vasanth :)	
18:41:42	From	Vasanth Ramani	to SUBJECT(Privately) :	ok
18:41:50	From	Vasanth Ramani	to SUBJECT(Privately) :	ok
18:42:33	From	SUBJECT	to Vasanth Ramani(Privately) :	sorry for sending it to the wrong audience
18:42:59	From	Vasanth Ramani	to SUBJECT(Privately) :	np man
18:43:18	From	SUBJECT	to Vasanth Ramani(Privately) :	::)
19:00:12	From	Vasanth Ramani	to SUBJECT(Privately) :	what are we discussing? I missed it.
19:00:24	From	Vasanth Ramani	to SUBJECT(Privately) :	what are we discussing? I missed it.

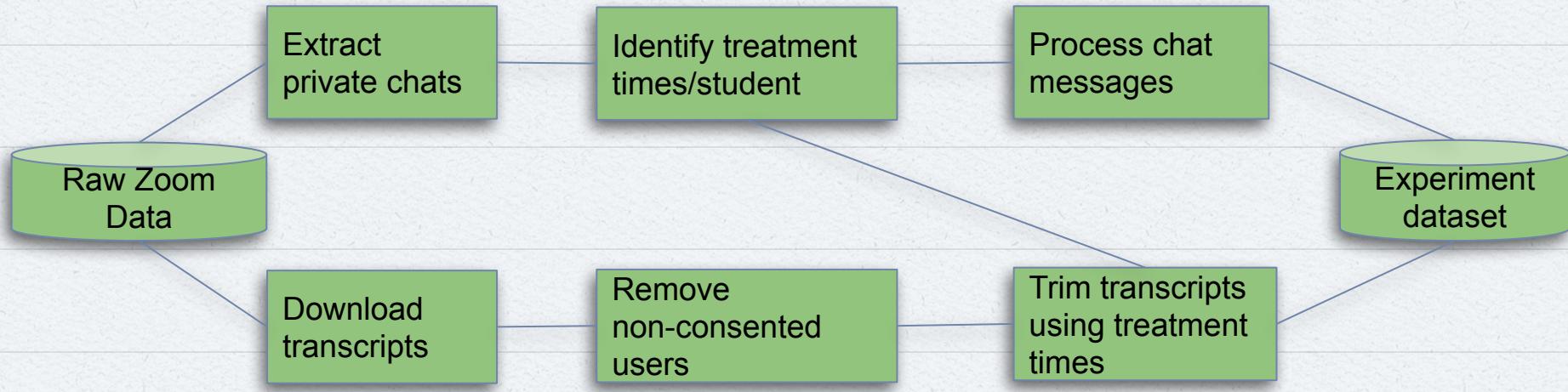
3

Section	Treatment
Pilot Sec4 10292020	Sorry what were we discussing? I missed it
Pilot Sec5 10292020	What are we discussing? I missed it..., what are we discussing? I missed it.
Study Sec1 11172020	What were we discussing? I missed it
Study Sec2 11182020	What are we discussing? I missed it.
Study Sec3 11182020	hey <SUBJECTNAME>, what are we discussing right now? I missed it.
Study Sec5 11052020	Hey <SUBJECTNAME> did you get a handle on the conditions under which it's ok to use naturally occurring experiments? i'm still not sure i understand...
Study Sec5 11192020	What are we talking about? I missed it., Hey <SUBJECTNAME>, do you know what were we discussing? I missed it !
Study Sec5 12032020	I just went to grab a snack. What paper are people talking about now?, was there a paper on this? or just the async info on it? I'm a little behind, I just went to grab a snack. What was the paper you sent?
Study w266 11032020	Hi. What are we discussing. I missed it.

# PROCESS FLOW



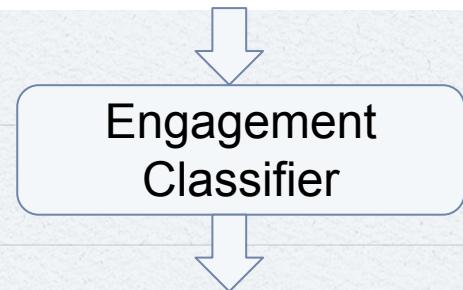
# TEXT/NLP PIPELINE



## TOOLS

- ✗ Zoom
- ✗ Python libraries
- ✗ Jupyter notebook
- ✗ Custom automation scripts
- ✗ R statistical packages

# MEASUREMENT GOALS



# LABELING DATASET FOR TRAINING AND TESTING

## Annotation Software Sample Image

606. Zoom Picture Scoring

Anonymized → 

**1. Select the person's behavioral appearance:**

On Task  
 Off Task  
 Can't decide

**2. Select the person's emotional appearance:**

Satisfied  
 Confused  
 Bored  
 Can't decide

*On Task:* The student is looking towards the screen or looking down to the keyboard below the screen.  
*Off Task:* The student is looking everywhere else or eyes completely closed, or head turned away.  
*Can't Decide:* If you cannot decide on the behavioral state.

*Satisfied:* The student is not having any emotional problems during the learning task. This can include all positive states of the student from being neutral to being excited during the learning task.  
*Confused:* The student is getting confused during the learning task. In some cases, this state might include some other negative states such as frustration.  
*Bored:* If the student is feeling bored during the learning task.  
*Can't Decide:* If you cannot decide on the behavioral state.

**Submit**

## Scoring Evaluation\*

Behavioral	Emotional	Engagement
On-task	Satisfied	Engaged
On-task	Confused	Engaged
On-task	Bored	Disengaged
Off-task	Satisfied	Disengaged
Off-task	Confused	Disengaged
Off-task	Bored	Disengaged

Flask Annotation App:  
<https://engagement241.herokuapp.com/>

\*Reference:

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., Picard, R.: Affect-aware tutors: recognising and responding to student affect. International Journal of Learning Technology



# MEASUREMENT OF AGREEMENT

## Student Engagement Dataset:

- 1,333 zoom images
- 3 MIDS classes (6 sections), 9 hours
- 5 student labelers
- Fleiss Kappa: 0.413

The following table can be used to interpret the IRR

Kappa	Interpretation
< 0	Poor Agreement
<b>0.01 – 0.20</b>	Slight Agreement
<b>0.21 – 0.40</b>	Fair Agreement
<b>0.41 – 0.60</b>	Moderate Agreement
<b>0.61 – 0.80</b>	Substantial Agreement
<b>0.81 – 1.00</b>	Almost Perfect Agreement

## Dataset Biases & Challenges:

- Engagement Imbalance
- Recognizing fellow students
  - Positions near camera
  - Learned facial expressions
- Unimodal labeling
  - Good for removing biases
  - Bad for model training

S.E.D.	Engaged	Disengaged
Training	576	173
Validation	257	77
Test	192	58
Totals	1025	308



# FRAME PROCESSING

9 Recorded Zoom Videos

1



Trim Videos with Python ffmpeg and moviepy

2



Sample gallery stills taken at 1 - 12 fps

3



Microsoft Optical Character Recognition (OCR), FuzzyMatch names, and Image Processing - Filter & Crop

4

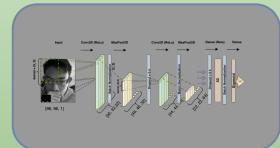


Face Cropping, Resizing Grayscale Feature Extraction, Normalize

5



CNN Classifier



36 pre-treatment and post treatment clips

25,263 gallery and individual stills taken from 60 subjects

[Subject X, N frames, Date, Section, ..]

[96 x 96]

## TOOLS

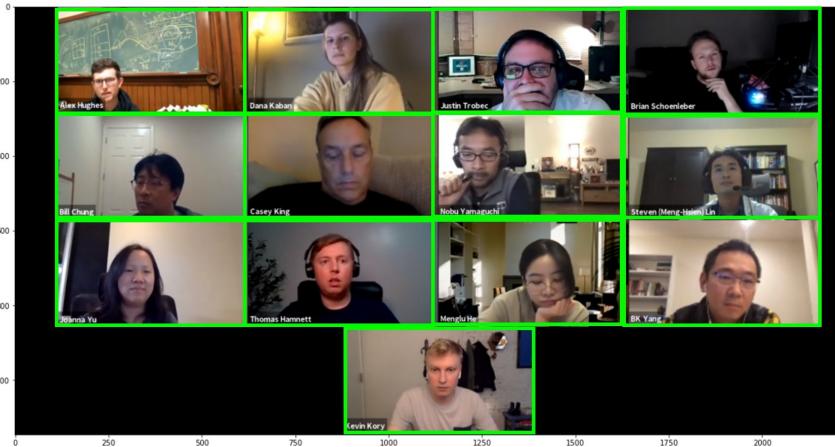
- ❑ Zoom
- ❑ OpenCV, Azure Computer Vision API
- ❑ Windows Photos
- ❑ Python libraries, Jupyter notebook
- ❑ Custom automation scripts
- ❑ R statistical packages



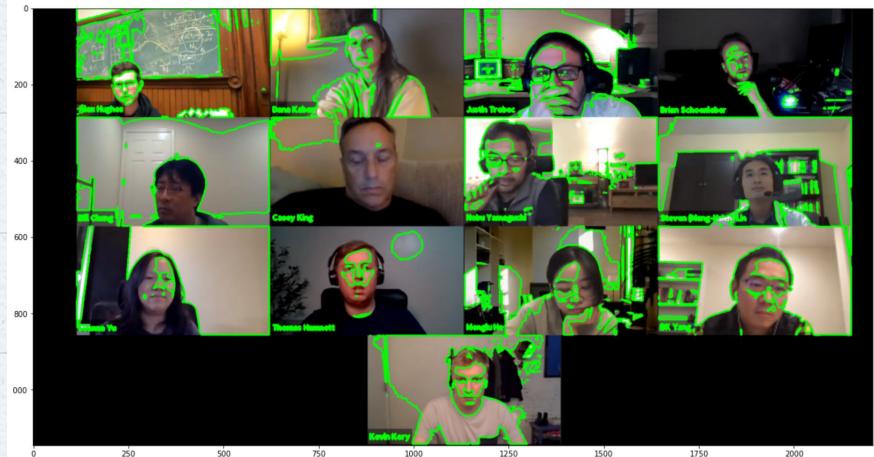
# CROPPING INDIVIDUAL PICTURES FROM A FRAME: ATTEMPT 1

Idea: can we use black contours around rectangles?

Expectation...



Reality :-)



# CROPPING INDIVIDUAL PICTURES FROM A FRAME: WHAT WORKED

1

```
{'boundingBox': '1187,1745,107,17',  
 'lines': [{  
     'boundingBox': '1187,1745,107,17',  
     'words': [{  
         'boundingBox': '1187,1745,74,17',  
         'text': 'Joanna'},  
        {'boundingBox': '1269,1745,25,17',  
         'text': 'Yu'}]}]}],
```

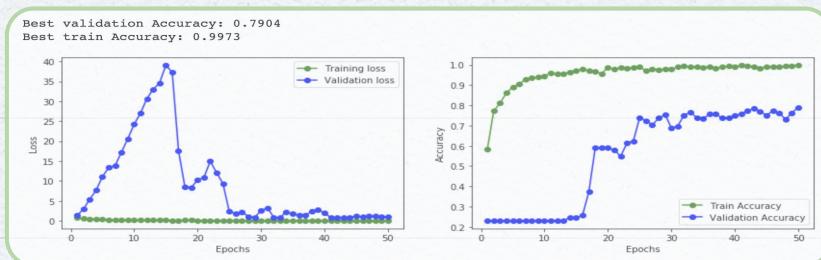
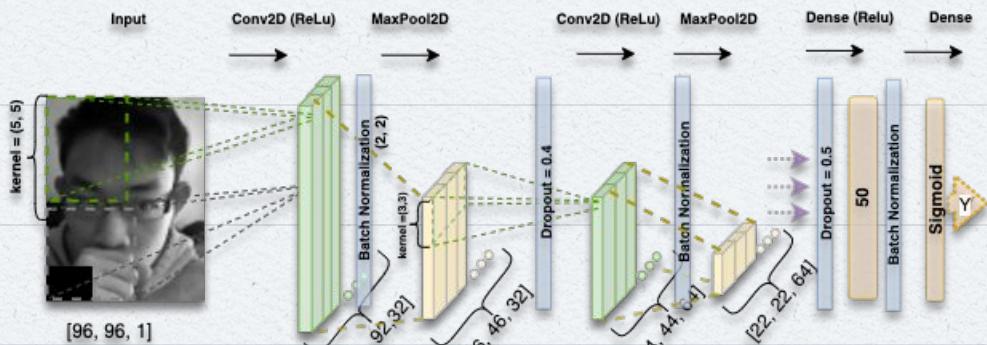
2

3

Changing Nobu Ya'naguchi to Nobu Yamaguchi  
Nobu Yamaguchi is already in the set  
Changing Dana Kabm to Dana Kaban  
Dana Kaban is already in the set  
Changing Justh Trobec to Justin Trobec  
Justin Trobec is already in the set  
Changing asey Khg to Casey King  
Casey King is already in the set  
Changing Bil Chung to Bill Chung  
Bill Chung is already in the set  
Changing Steven (Meng-Hsien)... to Steven (Meng-Hsien) Lin  
Steven (Meng-Hsien) Lin is already in the set  
Changing Kevh Kory to Kevin Kory  
Kevin Kory is already in the set

- ✗ Send the frame to OCR API
  - ✗ Get parsed texts
  - ✗ Get coordinates
- ✗ List participants from json response
  - ✗ Infer size & coordinates of images
  - ✗ Use fuzzy string matching to correct parsing
- ✗ Cut named sub-images

# MEASUREMENT MODEL



===== Evaluating model on the validation set  
Accuracy: 79.04 F1\_score: 68.70

	precision	recall	f1-score	support
0	0.55	0.47	0.51	77
1	0.85	0.89	0.87	257
accuracy	0.70	0.68	0.69	334
macro avg	0.70	0.68	0.69	334
weighted avg	0.78	0.79	0.78	334

===== Evaluating model on the test set  
Accuracy: 74.40 F1\_score: 65.30

	precision	recall	f1-score	support
0	0.45	0.50	0.48	58
1	0.84	0.82	0.83	192
accuracy	0.65	0.66	0.65	250
macro avg	0.65	0.66	0.65	250
weighted avg	0.75	0.74	0.75	250



# MEASUREMENT UNITS

Pre Treatment Frames



Score =  
sum( engaged )



2

Percentage Engaged (PE) =  
Score / total frames

$$2 / 5 = 0.4$$

Output =  
100 x diff (pre PE, post PE )



$$(0.75 - 0.4) \times 100 = 35$$

Post Treatment Frames



3\*

$$3 / 4 = 0.75$$



Text (2 channels)



\* We consider 10 sec of speaking as an equivalent to 'engaged' of a single frame



# 2. EXPERIMENTS

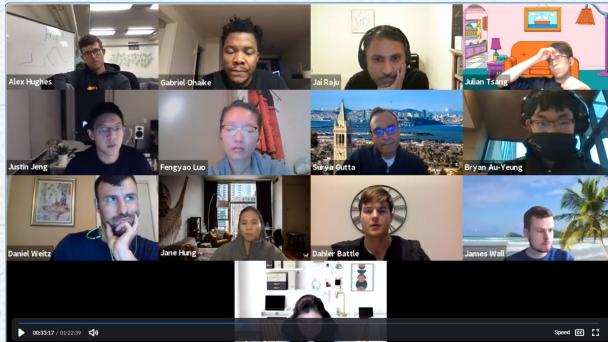


# EXECUTION OF EXPERIMENTS

## Administering Treatment

- Pilot Study Spillovers
- Disabled chat windows

**Novelty Effect:**  
Section 1 (Micah) with Alex substitution



## Compliance Issues

Compliance is defined as subjects who responded to our treatment and participated with their cameras on in class.

## Attrition

- Internet disconnection after treatment
- Impossible measurement (no audio, visual, or text response)
- Break out room during monitoring period

Applied Treatment	Chat Response	No Chat Response
-------------------	---------------	------------------

Video or Speech

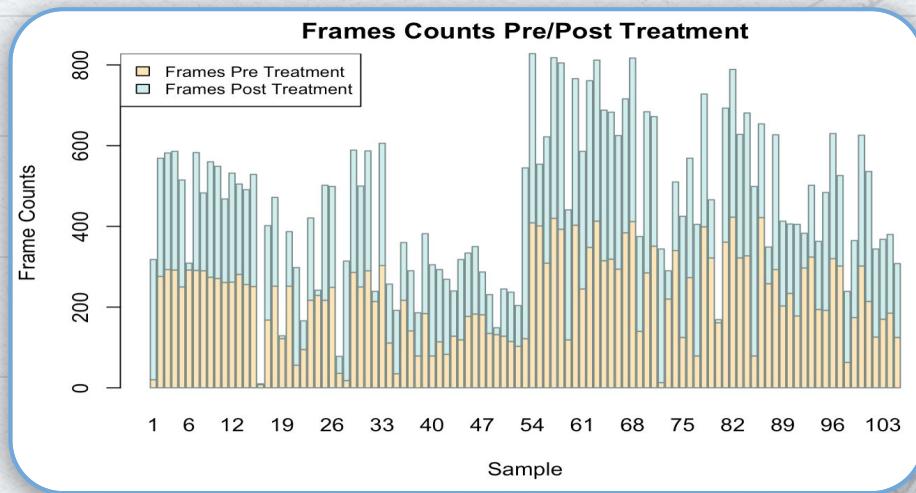
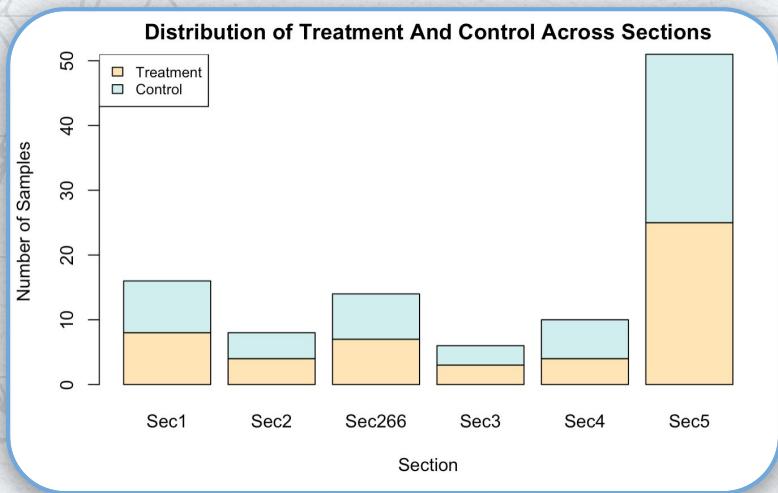
Compliers	Partial Compliers
Attrition / Partial TE	Attrition

No Video or Speech

Taking it further: Use chat response as engagement signals.



# EXPLORATORY DATA ANALYSIS



## Experiment Results

	Models			
	Treatment	Sec (B)	Sec+Date (B)	Sec+Date+Study (B)
Treatment	-2.15 (1.68)	-1.94 (1.73)	-1.92 (1.78)	-1.62 (1.94)
Section 2		-0.82 (7.05)		
Section 266		1.08 (2.15)		
Section 3		2.47 (2.72)		
Section 4		<b>-6.46** (3.28)</b>		
Section 5		-0.81 (2.12)		
10/29/2020 : Section 5			<b>7.02** (3.27)</b>	
11/03/2020 : Section 266			<b>7.54** (2.99)</b>	
11/05/2020 : Section 5			4.81 (3.68)	
11/17/2020 : Section 1			<b>6.46** (3.28)</b>	
11/18/2020 : Section 2			5.64 (7.29)	
11/18/2020 : Section 3			<b>8.93*** (3.42)</b>	
11/19/2020 : Section 5			5.75 (4.49)	
12/03/2020 : Section 5			5.19 (3.52)	
10/29/2020 : Section 4 : Study 2				-4.14 (7.88)
10/29/2020 : Section 5 : Study 1				3.08 (2.01)
10/29/2020 : Section 5 : Study 2				<b>7.70** (3.59)</b>
11/03/2020 : Section 266 : Study 1				<b>5.21** (2.53)</b>
11/03/2020 : Section 266 : Study 2				<b>6.63*** (2.09)</b>
11/05/2020 : Section 5 : Study 1				-1.14 (3.31)
11/05/2020 : Section 5 : Study 2				<b>8.03** (3.80)</b>
11/17/2020 : Section 1 : Study 1				5.15 (3.33)
11/17/2020 : Section 1 : Study 2				4.52 (2.86)
11/18/2020 : Section 2 : Study 1				2.29 (14.03)
11/18/2020 : Section 2 : Study 2				5.73 (7.58)
11/18/2020 : Section 3 : Study 1				<b>8.88* (4.56)</b>
11/18/2020 : Section 3 : Study 2				<b>5.72** (2.51)</b>
11/19/2020 : Section 5 : Study 1				4.12 (3.85)
12/03/2020 : Section 5 : Study 1				5.39 (4.36)
12/03/2020 : Section 5 : Study 2				1.73 (2.66)
Constant	1.47 (1.22)	2.14 (2.07)	- 4.33 (2.89)	-2.85 (1.84)
Observations	105	105	105	105
R2	0.02	0.07	0.08	0.15
Adjusted R2	0.01	0.02	-0.01	-0.01

Symbols:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01  
(B) = Blocking, Sec = Section

# RESULTS

- Positive effect in most sections, blocked by study & date
  - More than half - statistically significant!
  - More covariates can be included in future studies
- We observed that subjects are impacted in different ways
- RSEs don't account for classifier's accuracy
- No significant anova F test results
- Spillover between studies
- Priming effect in some sections

3.

# CONCLUSION



# OUR CONTRIBUTIONS

- ✗ A prototype of system for automated evaluation of experiments with online learning
  - ✗ Methodology, including design and automated analysis of experiments
  - ✗ Data collection pipeline, leveraging computer vision and NLP AI for extraction of signals
  - ✗ DNN for classification of engaged/disengaged with high accuracy
  - ✗ Labeled dataset for training and validation of the model
- ✗ Proof-of-concept application of the system in our class project
  - ✗ Conducted 7 controlled experiments on w241 students across 5 sections
  - ✗ Positive effect of the treatment detected in most sessions with statistical significance
  - ✗ Rejected null hypothesis that private chat discussions are having no impact on engagement



# FUTURE WORK

- ✗ Get Berkeley approval for external publication of the research results
- ✗ Improve the performance of engagement prediction
  - ✗ More labeled samples for the training/validation dataset
  - ✗ Interpretability improvements and new signals
  - ✗ Sliding window labeling of images series for training the model
- ✗ Integrate with interactive educational platforms
  - ✗ Data collection pipeline reliability improvements
  - ✗ Improvement of name detection with privacy preserving
  - ✗ More advanced NLP methods for sentiment detection
- ✗ Report detected issues with Cognitive Services API
  - ✗ 400 HTTP status response on certain valid images
  - ✗ OCR failing on some images where the text is clearly visible



# REFERENCES

- ✖ P. Goldberg, Ö. Sümer, K. Stürmer, W. Wagner, R. Göllner, P. Gerjets, E. Kasneci & U.Trautwein. "Attentive or Not? Toward a Machine Learning Approach to Assessing Students' Visible Engagement in Classroom Instruction" (2019). Educational Psychology Review URL: <https://link.springer.com/article/10.1007/s10648-019-09514-z>
- ✖ O. Mohamad Nezami , M. Dras, L. Hamey, D. Richards, S. Wan, & C. Pari (2018). "Automated Recognition of Student Engagement using Deep Learning and Facial Expression" (2019) URL: <https://arxiv.org/abs/1808.02324>
- ✖ J. Whitehill, Z. Serpell, Y. Lin, A. Foster and J. R. Movellan (2014) "The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions," in IEEE Transactions on Affective Computing, vol. 5, no. 1, pp. 86-98, 1 Jan.-March 2014, doi: 10.1109/TAFFC.2014.2316163. URL: <https://inc.ucsd.edu/mplab/wordpress/wp-content/uploads/EngagementRecognitionFinal.pdf>
- ✖ J. Costa., M.. Jung, M. Czerwinski, F.Guimbretière, T. Le and T. Choudhury (2018). "Regulating Feelings During Interpersonal Conflicts by Changing Voice Self-perception." CHI '18 URL: <https://www.microsoft.com/en-us/research/uploads/prod/2018/03/CHI2018-conflicts-final.pdf>
- ✖ S. Mota and R.W. Picard (2003),"Automated Posture Analysis for Detecting Learner's Interest Level,"Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, CVPR HCI, June, 2003. PDFTR 574. URL: <https://affect.media.mit.edu/pdfs/03.mota-picard.pdf>

# Q&A

