Using AI to theme 10,000 survey comments in less than 1 hour

Wayde Gilliam, Sr. Software Architect/Engineer
e: wgilliam@ucsd.edu

Angela Song, Sr. Director, Operational Strategic Initiatives
e: aysong@ucsd.edu

NCCI 2021

VIRTUAL ANNUAL CONFERENCE | JULY 21–22

Introductions

Goals for this session

1. UC San Diego's Tritonlytics Program

2. Demystifying the AI/ML jargon

3. The capabilities and limitations of ML in Natural Language Processing (NLP)

4. An example: A clustering/summarization ML pipeline for theme discovery

5. How to get started with AI in your organization

36 Surveys

17 Institutions

1,833 Departments

268,144 Total Invites

4 University Systems

3 States

2 Non-profit Organizations

85,994 Survey Comments

766,832 Survey Comments (Total)

Demystifying the AI/ML jargon

Artificial intelligence v. machine learning v. deep learning

How does it work?

When someone uses 'Machine learning', 'AI' and 'deep learning' interchangeably in a discussion
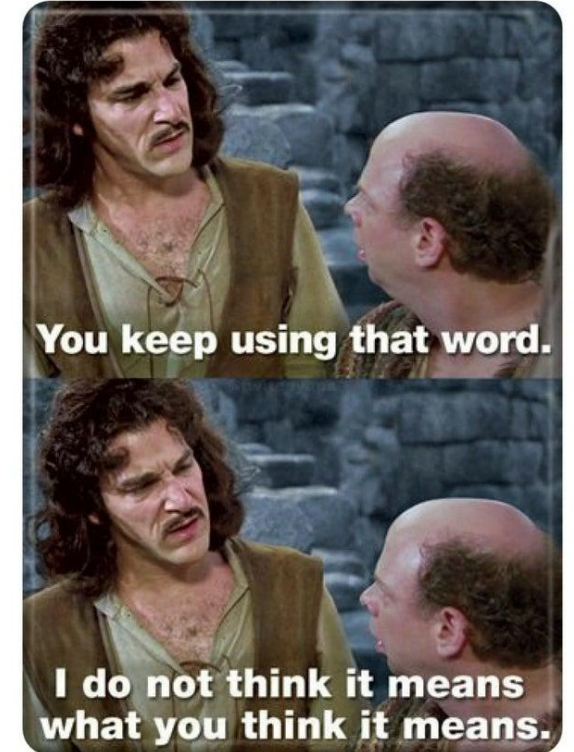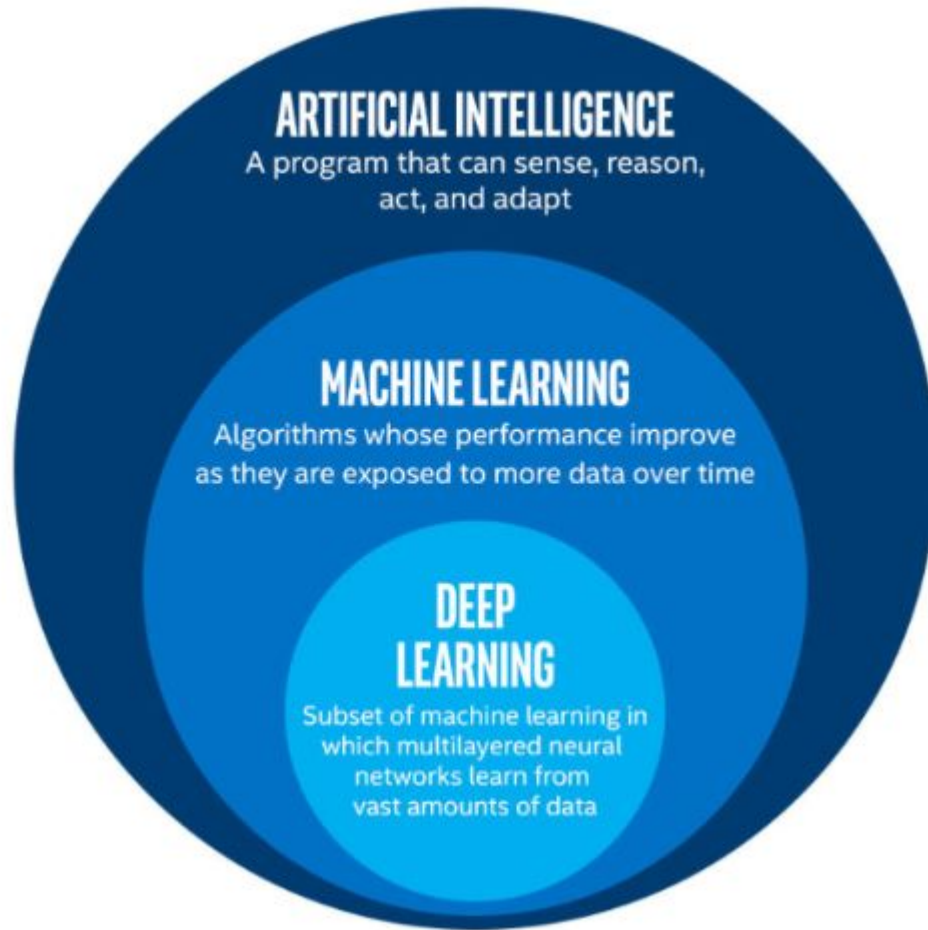
Demystifying the AI/ML jargon: **What is it?**

ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt

MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time

DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data

Artificial intelligence is … "a science [that] studies ways to build intelligent programs and machines that can creatively solve problems, which has always been considered a human prerogative"

Machine learning is … "a subset of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed" using traditional algorithms like linear or logistic regression.

Deep learning is … "a subset of machine learning, which utilizes neural networks"

Demystifying the AI/ML jargon: **What is it?**

Traditional programming

```
balance = 1000

def can_withdrawl(amt):
    return amt <= balance

print(f'Can I withdrawl a $100 bucks? {can_withdrawl(100)}')
print(f'Can I withdrawl a $1000 bucks? {can_withdrawl(1000)}')
print(f'Can I withdrawl a $10000 bucks? {can_withdrawl(10000)}')
```

```
Can I withdrawl a $100 bucks? True
Can I withdrawl a $1000 bucks? True
Can I withdrawl a $10000 bucks? False
```

Machine learning

$Y = a + bx$

Deep learning
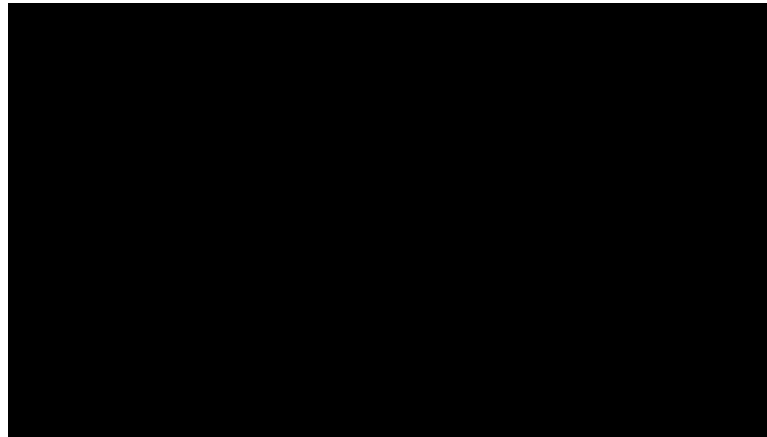
"… *a computational framework for modeling cognitive process* that seems … closer than other frameworks to the style of computation as it might be done by the brain" … based on how real neurons work.

From 'Deep Learning for Coders with fastai & PyTorch' ch.1

Animation (ML): https://towardsdatascience.com/linear-regression-5100fe32993a
Animation (DL): https://pub.towardsai.net/what-is-machine-learning-ml-b58162f97ec7?sk=aa068bdef0ec674e9fc95c7d9698e844&source=friends_link&gi=5245c2141068

How folks imagine it works ...
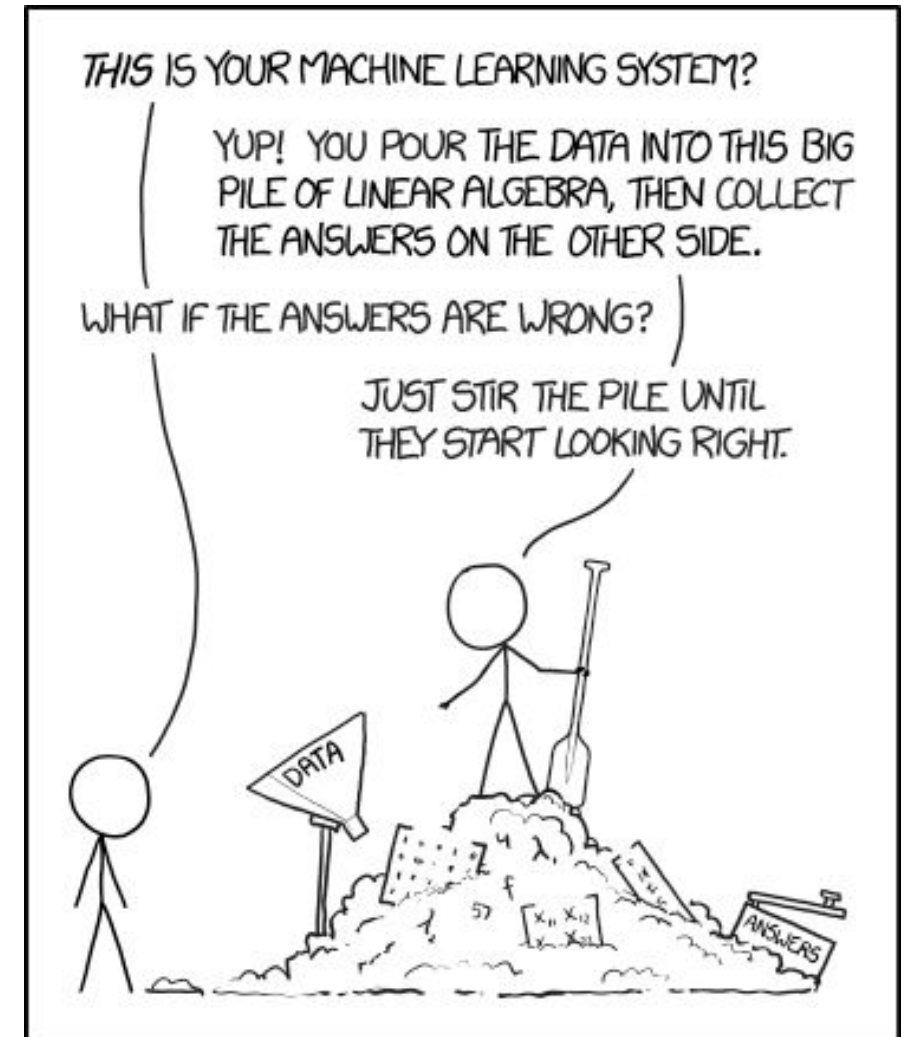
Photo: xkcd.com

Demystifying the AI/ML jargon: **How does it work?**

How it *actually* works …

The "data not including the labels" also called the *independent variables* (e.g., survey comments, or images of dogs and cats)

The *functional form of the model* which is a "special kind of program: it's one that can do many different things" depending on the *parameters*

The correct "labels" associated with each of our "inputs" (known also as "targets" or the "dependent variable"

The *variables that can be updated* to improve the performance of the model

The *results of passing our independent variables (our "inputs") through the model*

The *measure of performance* … smaller is better

From 'Deep Learning for Coders with fastai & PyTorch' ch.1

Capabilities and limitations of ML

Where does machine learning work well in natural language processing (NLP)?

What are the limitations of of machine learning in NLP?

What can we do to mitigate these limitations?

Capabilities and limitations: **Where is ML useful?**

With regards to NLP, machine learning is increasingly becoming the go-to solution for the following tasks:

			Summarization	Translation
Sequence classification	**Token classification**	**Question Answering**	**Text generation**	
Language modeling				
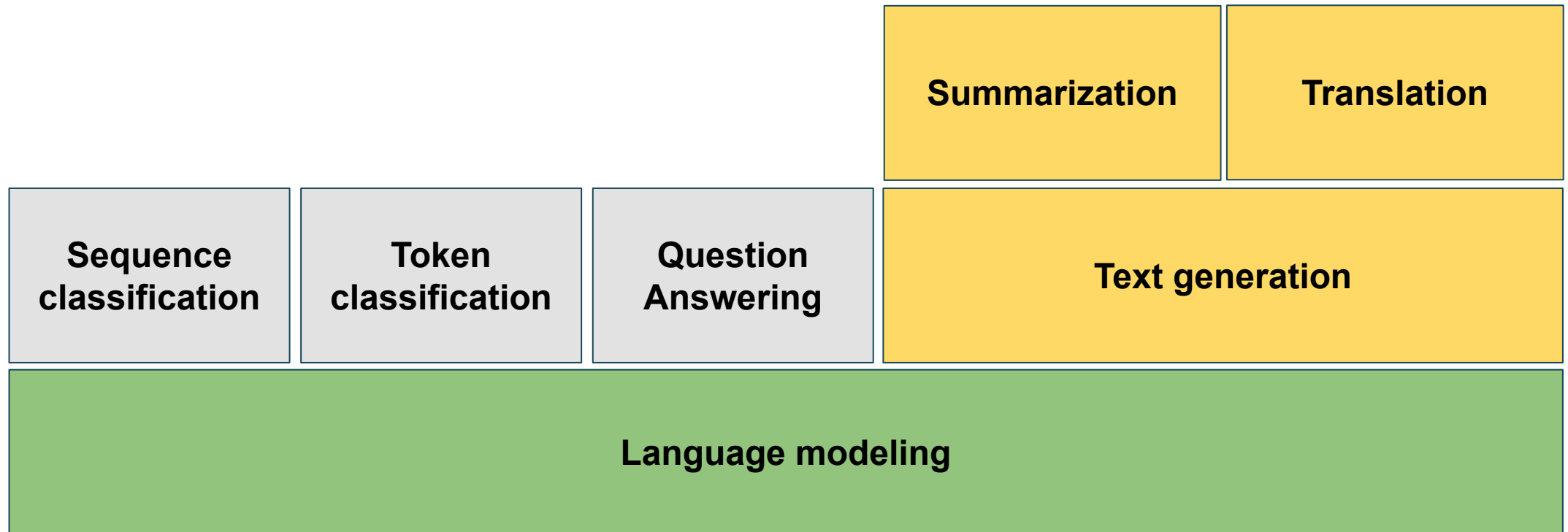
Capabilities and limitations: **What are its limitations?**

ML isn't perfect however. Here are some things to think about when using it …

1. **ML isn't perfect (I think I said that already ;-))**

"I am a senior advisor at the *Association of Northern California Counties (ANR)* and I …"

DOES NOT EXIST!!!

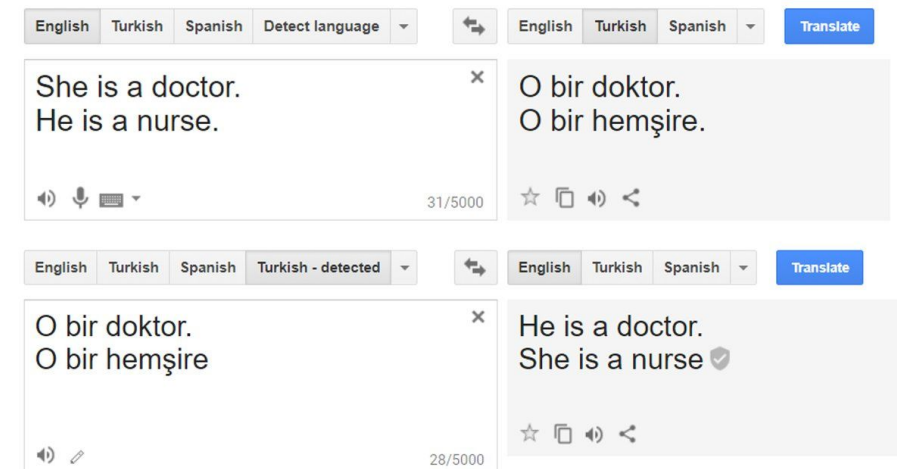2. **Models are only as good as the data they are trained on**

"This is just the beginning, wait till apes come back from moon. We will help more apes , and I wanna help a Rhino. Because apes together strong as many 🦏"

Courtesy of reddit user zombieattakc

3. **Bias inherent in your training data will be reflected in your model's predictions**

Capabilities and limitations: **Mitigating the limitations**

Limitation	Mitigation
1. **ML isn't perfect (I think I said that already ;-))**	- Manage your expectations and objectives - Define an acceptable measure of performance first *Augment your ML predictions with human review*
2. **Models are only as good as the data they are trained on**	- Ensure models are trained on data similar to what you plan to use - Fine-tune your models trained on dissimilar data with your own labeled data *Augment your ML predictions with human review*
3. **Bias inherent in your training data will be reflected in your model's predictions**	- Understand what data the models you are using were trained on - Be aware of stated biases (real or potential) by dataset authors *Augment your ML predictions with human review*

Demo: Using AI for theme discovery

How it works: The Augmented ML/Human clustering/summarization pipeline

Exploring the results

Using AI for theme discovery: **The ML pipeline**

1. Build a numerical representation of each sentence *using a DL model trained on a sentence similarity task*

2. Group texts *using a ML clustering algorithm* so that related texts are grouped together and related groups are close to one another.

3. Reduce that numerical representation's dimensionality to two so it can be visualized *using a ML dimensionality reduction algorithm*.

4. Based on the probability a sentence is associated with the found cluster and also its centrality in it, rank them from most exemplar to least

5. *Apply a DL model trained to summarize* each group (or cluster) as a topic (or theme)

6. *Augment final results with human reviewers* who will correct theme predictions, combine discovered themes they feel are the same, and also remove sentences associated with a theme that they believe are not relevant

7. Use reviewed/corrected data to *improve all ML* models periodically

Using AI for theme discovery: **DEMO**

This all sounds fine and dandy, but *how can we present such findings so that they are meaningful and actionable?*
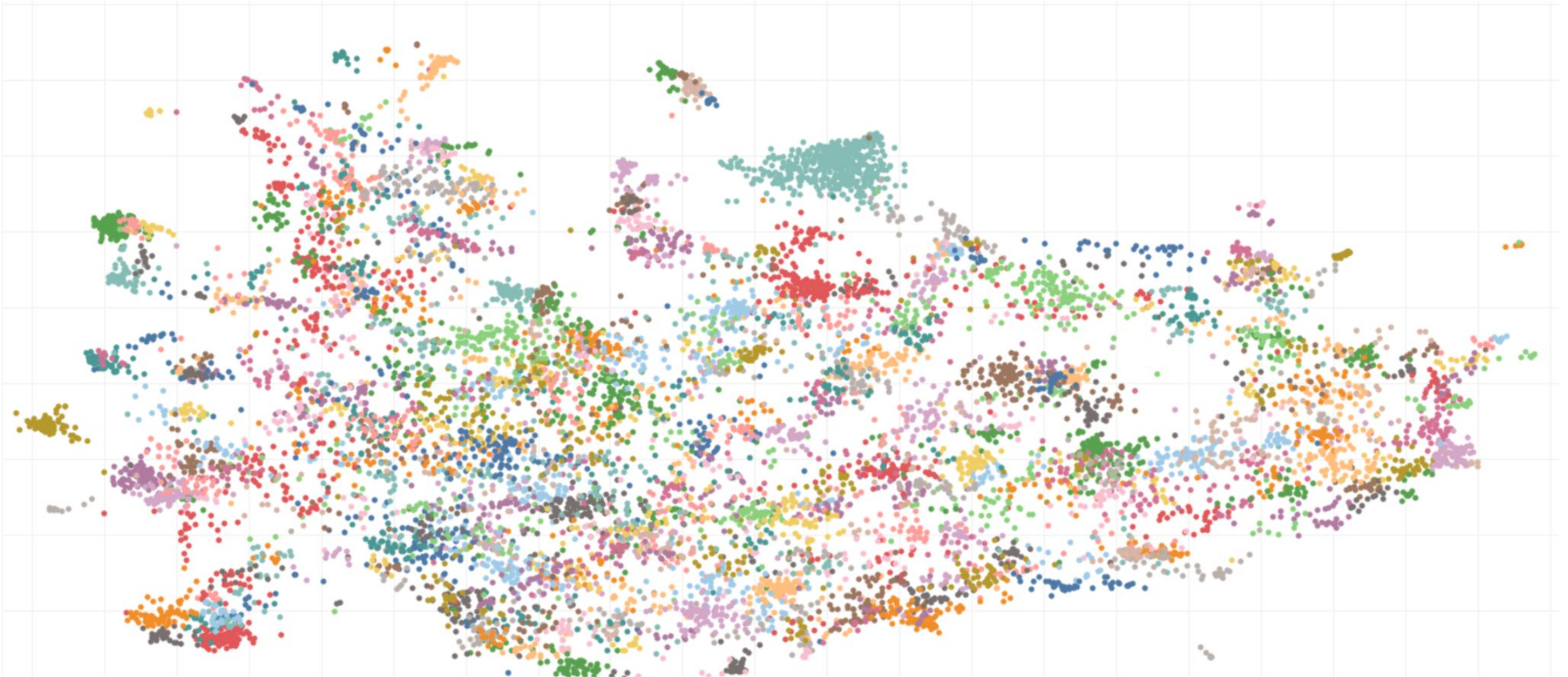
AI Applications and Examples:

1) *SURVEYS*
   a) COVID-19 Student Risk and Needs Survey

2) *STRATEGIC PLANNING*
   a) Teaching and Learning Commons' Stakeholder Interview Synthesis

3) *CROWD-SOURCING*
   a) IdeaWave Campaign - Reimagining the Workplace

COVID-19 Student Risks and Needs Survey

Over **650 comments** to review and make sense of

~~16 hours~~ → **15 minutes, identified top themes and their exemplar sentences**

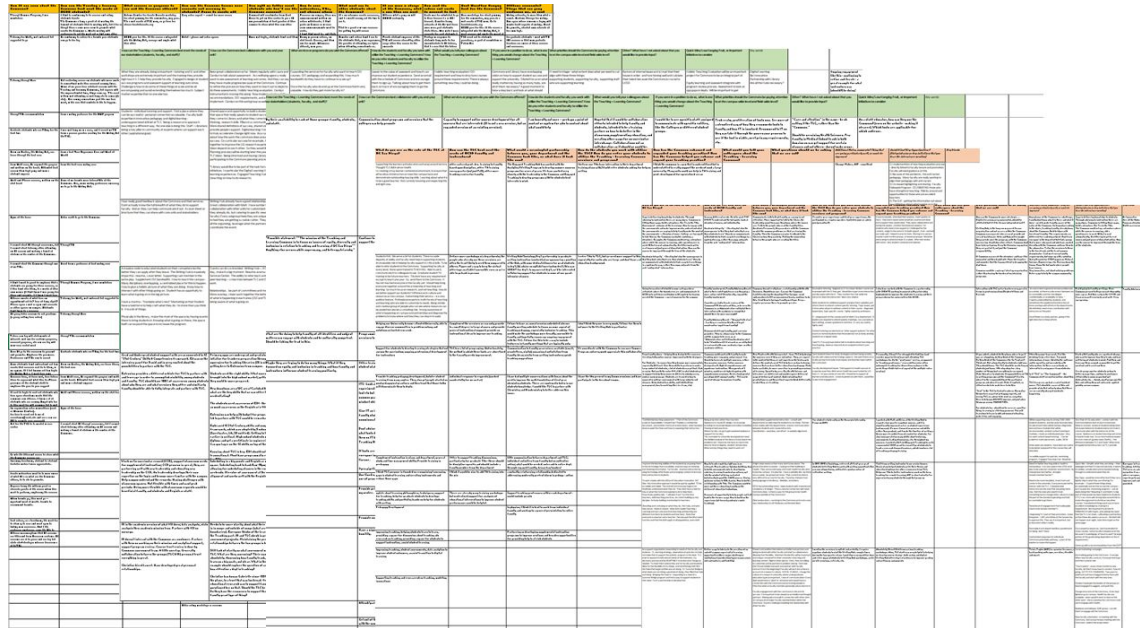| Augmented ML Theme	Raw ML Cluster	# of Sentences	Examplar Sentences
More Strict Rules Enforcement	MORE SUPERVISION FOR CAMPUS PROTOCOLS	210	To encourage everyone to wear mask covering up to nose / A lot of people who visit the campus do not wear masks properly/Enforce Social Distancing Rules/Strictly enforce rules and guidelines.
Don't Rush In-Person Classes	NOT RUSH IN PERSON CLASSES	180	Please don't do in person classes for anyone/There is no need to begin in-person classes until there is a vaccine available/I would not want to be on campus attending in person class
Continue Providing Asymptomatic Testing	ENCOURAGE REGULAR TESTING	173	More frequent testing/Continue offering COVID-19 tests/More Testing Locations (Esp. Drive-Thru)/Testing Options/Provide Testing for Housemates&Family
Satisfaction with RTL	THANK YOU/YOU ALL ARE DOING GREAT/ KEEPING US SAFE THANKS FOR/UCSD IS DOING THE BEST	160	I think you've been doing a wonderful job and I feel safe on campus/Thank you for your efforts / You're doing a great job / Thank you for keeping us safe / I think UCSD is doing a good job / I'm impressed with the efforts that UCSD is making to keep students stay healthy and safe / Thank you! / I think they're doing a good job with what's going on now.
Dissatisfaction with Remote Learning	MANY STUDENTS FEEL ISOLATED ANDONLINE LEARNING IS HARD	94	Please do some more in person classes, it is very hard to feel social i'm any way on campus right now, many students feel isolated and it's very hard/online learning is very difficult/I feel the quality of online classes is much lower than in person classes/More in person classes can be done safely and will greatly increase learning outcomes.
No Roommates	STABLE HOUSING NO DOUBLE ROOMS	73	Don't add roommates to the double rooms that are currently acting as single rooms for the winter/spring quarter.
Mental Health Concerns	MENTAL HEALTH IS MAJOR CONCERN	55	Help people with mental health first/I also fear isolation because I don't want my anxiety to worsen/burned out/feeling stressed
Provide PPE	CAMPUS MORE HAND SANITIZER IN	46	Add more hand sanitizer dispensers in buildings/Provide masks and other sanitary items.
Concerns About Returning to Campus	NOT RETURNING TO CAMPUS	44	I disagree with having students on campus for any non-essential reasons/Do not move more people onto campus.
Provide Finanical Support	MORE FUNDS TO SUPPORT STUDENTS	40	Financial support during this time is needed for students/Reduce Activity Fees/Provide Rent Subsidies and Stipends
Food Options	CAMPUS FOOD PANTRY AND DISTRIBUTION	34	Provide free food to students who live on-campus (grad housing) to limit going out grocery shopping (aside from the food pantry and food recovery network)
Open More Study Spaces	PLEASE OPEN MORE STUDY SPACES	26	please open some study space or reduce tuition accordingly.
Anxiety Caused by COVID Focus	AND ANXIETY CAUSED BY COVID	26	I simultaneously want to avoid all person-to-person interactions due to deep fear of contracting COVID, while feeling the emotional stress of not having had positive social interaction for months.
Student Irresponsibility	FORCING STUDENTS TO RISK THEIR	25	A lot of students have been irresponsible with coronavirus off and on campus, which is worrisome to people who want to protect their health at all costs and are forced to live with people who don't feel the same way.
Safety Concerns with Price Center Testing	CENTER TESTING FEELS REALLY UNSAFE	23	Price center testing feels really unsafe to me- 10 people crowded around a table without masks indoors terrifies me and makes me feel reluctant to take a test.
Professors Causing Stress and Busy Work	TOO MANY PROFESSORS ARE PRESSURING	23	Teachers have been regularly assigning lectures that exceed the allotted class time, creating a stressful and unnecessary surplus of work.
Problems with Locked Doors	PROBLEM WITH LOCKING THE DOORS	19	I should be able to use the same elevator as my apartment mate.
Socialization Options	WAYS FOR US TO SOCIALIZE	18	Look for ways to let people be social together, more than just seminars.

Crowd-Sourcing IDEAWAVE CAMPAIGN : Ways to meet diverse operational needs in a reimagined workplace

Theme	Example	Number of Instances
Respect for working hours.	"We respect working hours, and don't schedule meetings or expect work to occur outside of normal business hours."	123
More flexibility and options for how work is performed.	"For those who want to meet in person, or do not have a suitable place to work at home, clearly there must be in-person and hybrid work models as well as flex space available on all of the university's owned and leased properties throughout the region."	55
Commuting burden lessened.	"I worked standard work hours, and at the end of the day, I spend 15 minutes walking to my car and between 1-1.5 hours in traffic commuting from work --> preschool-->home. On a rainy day it was a 2 hour commute from work --> preschool-->home. I barely had enough time to get dinner on the table, and I felt like I was struggling to keep everything together."	52
Productivity and accountability are unique to each employee.	"You can tell when someone is not producing as the output is not what you would expect. If that is the case, you have a larger problem that will not be solved simply by being in the office."	44
High employee satisfaction.	"We can't go back to this [pre-pandemic work inflexibility]. It was not sustainable."	39

FROM THIS → **TO THIS**

Strategic Planning Interviews

150 stakeholder interviews

✓ Most mentioned themes from **Faculty**

predicted_theme	# mentions
MORE COMMUNICATION/ENGAGEMENT WITH FACULTY	33
INCREASE AWARENESS OF TLC	17
INSTRUCTION RECOMMENDATIONS	15
SUPPORT TEACHING/LEARNING STRATEGIES	14
COURSE DESIGN AND TEACHING SKILLS	13
WORK WITH FACULTY TO ENHANCE TEACHING QUALITY	10
MORE FACULTY AND TA TRAINING	9
ASSIGN AND USE SI	9
HELP FACULTY BE BETTER TEACHERS	9
ROLE OF COMMONS	9

✓ Most mentioned themes from **Students**

predicted_theme	# mentions
FROM THE CLASSROOM AND TA'S	23
CLARIFY SERVICES, WORKSHOPS, MENTOR TRAINING	16
CONTINUE VIRUTAL OPTIONS	15
MAKE IT MORE GRAD STUDENT FRIENDLY	12
ENCOURAGE ITS USE BY PROFESSORS AND PEERS	12
ONGOING TUTOR TRAINING AND REVIEWS	12

✓ Most mentioned themes from **Campus Staff**

predicted_theme	_sentence
THEY USE THE SUPPLEMENTAL INSTRUCTION & WRITING HUB, THE ENGAGED	50
TLC HELPS FACULTY AND STUDENTS	43
COLLABORATING WITH OTHER DEPTS	32
MORE VISIBILITY, TRANSPARENCY AND BRANDING	28
DEVELOP ONLINE COURSES	19
PARTNERSHIPS	18
HELP FACULTY BETTER SUPPORT STUDENTS	16
MORE COMMUNICATION AND COLLABORATION BETWEEN HUBS	15

FROM THIS → **TO THIS**

Using AI in your own organization

Getting started with ML and *Tritonlytics*

Learning resources for the technical and technically challenged

Using AI in your organization: **Getting started**

The *Tritonlytics* team provides a variety of ML based solutions, from sentiment classification models, to named entity recognition (NER), summarization, and ad-hoc theme discovery. If you're interested in learning more, please feel out the interest form provided in the link below!
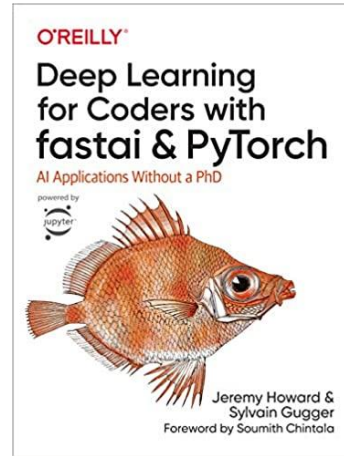
https://forms.gle/VR6ckixjsvk1ZvpX9

Using AI in your organization: **Learning resources**

Resources to get you on the path to understanding and using ML ...

Deep Learning for Coders with fastai & Pytorch
by Jeremy Howard & Sylvain Gugger

A great resource for project managers, those concerned with data ethics, and folks interested in developing ML models themselves

But what is a neural network? … a great explanation of what a NN is and how they work
by 3Blue1Brown (YouTube)

Twitter … the best platform for the latest in ML research and application

Questions?

Thank you!

To find out more about our services at UC San Diego, visit us at https://tritonlytics.ucsd.edu/

Wayde Gilliam, Sr. Software Architect/Engineer
e: wgilliam@ucsd.edu

Angela Song, Sr. Director, Operation Strategic Initiatives
e: aysong@ucsd.edu