What's so Hard about Natural Language Understanding?

Alan Ritter
Computer Science and Engineering
The Ohio State University

Collaborators:
Jiwei Li, Dan Jurafsky (Stanford)
Bill Dolan, Michel Galley, Jianfeng Gao (Microsoft Research), Colin Cherry (Google)
Jeniya Tabassum, Alexander Konovalov, Wei Xu, Marie Catherine de Marneffe (Ohio State)
Brendan O'Connor (Umass Amherst)

THE OHIO STATE UNIVERSITY

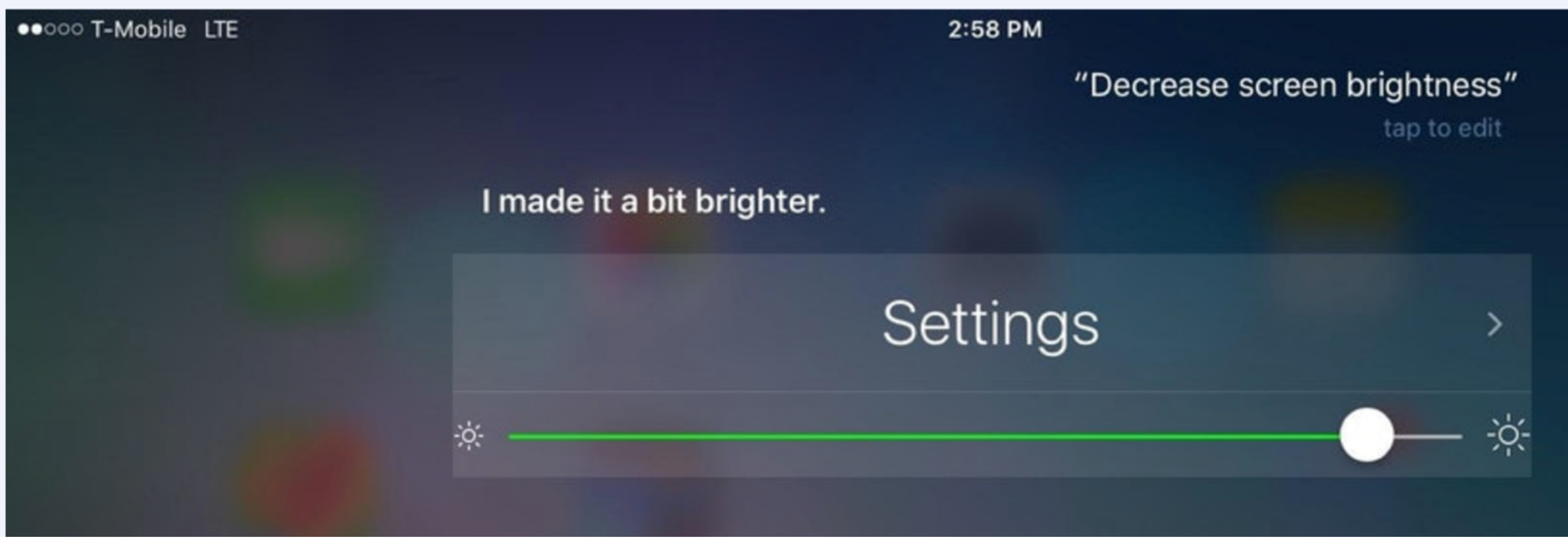
reddit /R/SIRIFAIL

48

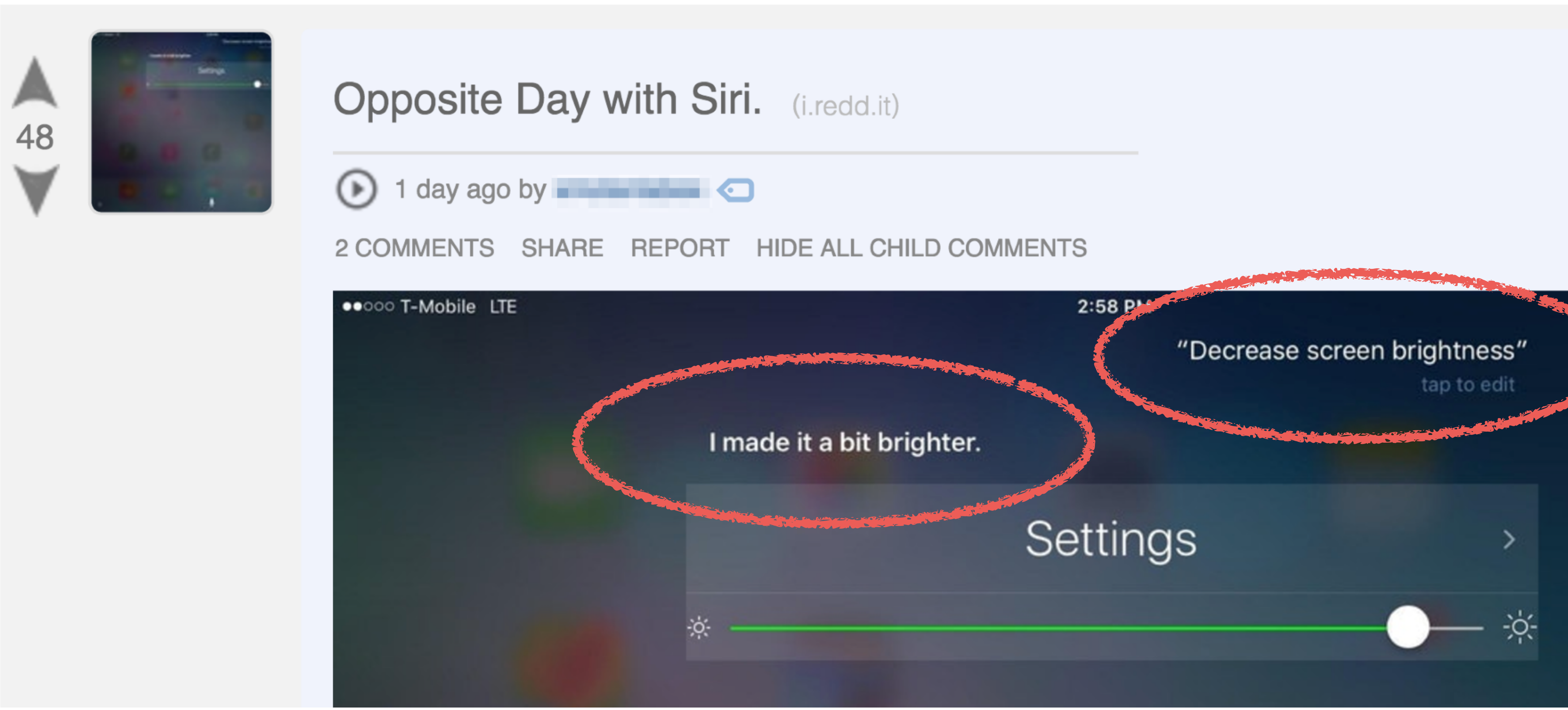
Opposite Day with Siri. (i.redd.it)

1 day ago by

2 COMMENTS SHARE REPORT HIDE ALL CHILD COMMENTS



reddit /R/SIRIFAIL



Q: Why are we so good at Speech, MT (but bad at NLU)?

People ***naturally*** translate and transcribe.

Q: Why are we so good at Speech, MT (but bad at NLU)?

People ***naturally*** translate and transcribe.

Q: Large, End-to-End Datasets for NLU?

- Web-scale Conversations?
- Web-scale Structured Data?

Q: Why are we so good at Speech, MT (but bad at NLU)?

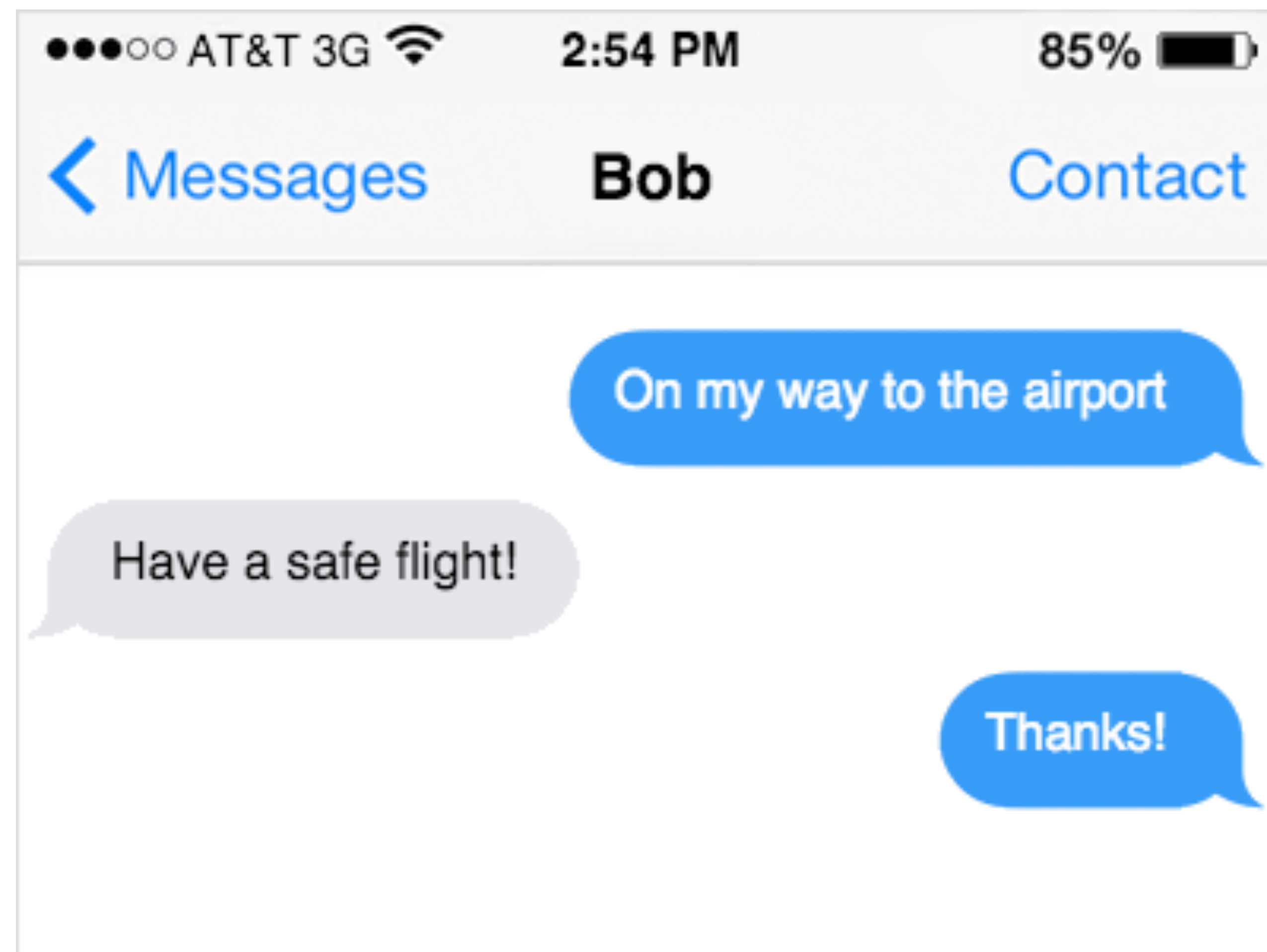
People ***naturally*** translate and transcribe.

Q: Large, End-to-End Datasets for NLU?

- Web-scale Conversations?
- Web-scale Structured Data?

Data-Driven Conversation

- **Twitter:** ~ 500 Million Public SMS-Style Conversations ***per Month***
- **Goal:** Learn conversational agents directly from massive volumes of data.



Data-Driven Conversation

- **Twitter:** ~ 500 Million Public SMS-Style Conversations ***per Month***
- **Goal:** Learn conversational agents directly from massive volumes of data.

[Ritter, Cherry, Dolan EMNLP 2011]

Noisy Channel Model

Input:

Who wants to come over for dinner tomorrow?

[Ritter, Cherry, Dolan EMNLP 2011]

Noisy Channel Model

Input:

$$\textbf{Who wants to come over } \underbrace{\textbf{for dinner}} \textbf{ tomorrow?}$$

Output:

Yum ! I

[Ritter, Cherry, Dolan EMNLP 2011]

Noisy Channel Model

Input:

Who wants to come over for dinner tomorrow?

Output:

Yum ! I want to

[Ritter, Cherry, Dolan EMNLP 2011]

Noisy Channel Model

Input:

Who wants to come over for dinner tomorrow?

Output:

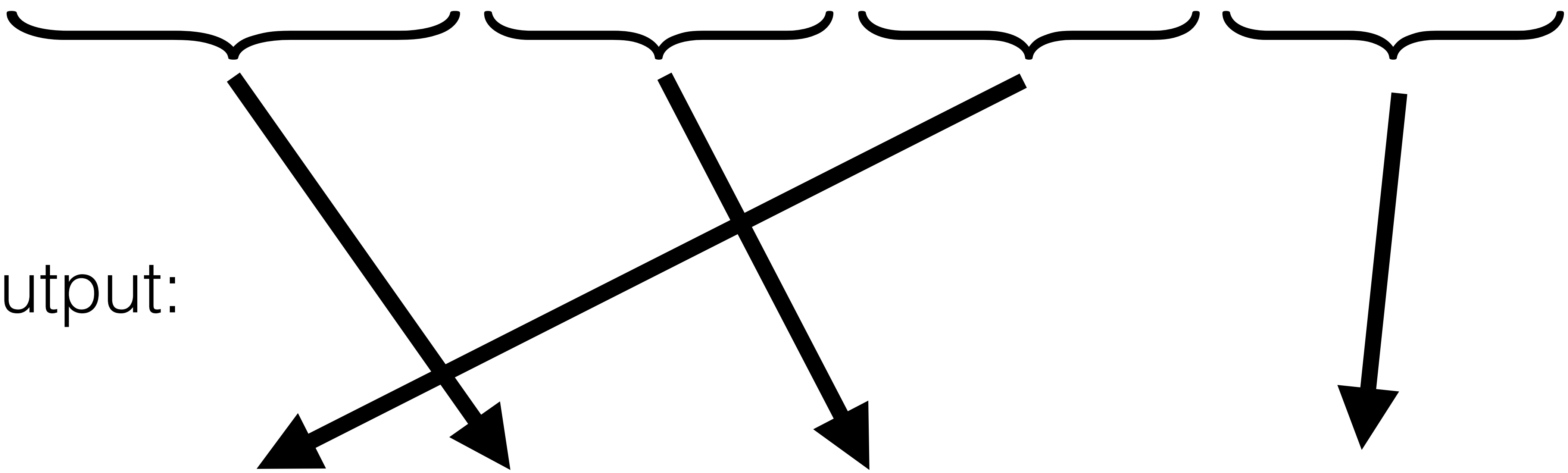
Yum ! I   want to   be there

[Ritter, Cherry, Dolan EMNLP 2011]

Noisy Channel Model

Input:

Who wants to come over for dinner tomorrow?



Output:

Yum ! I   want to   be there   tomorrow !

Neural Conversation

[Sordoni et. al. 2015] [Xu et. al. 2016] [Wen et. al. 2016]

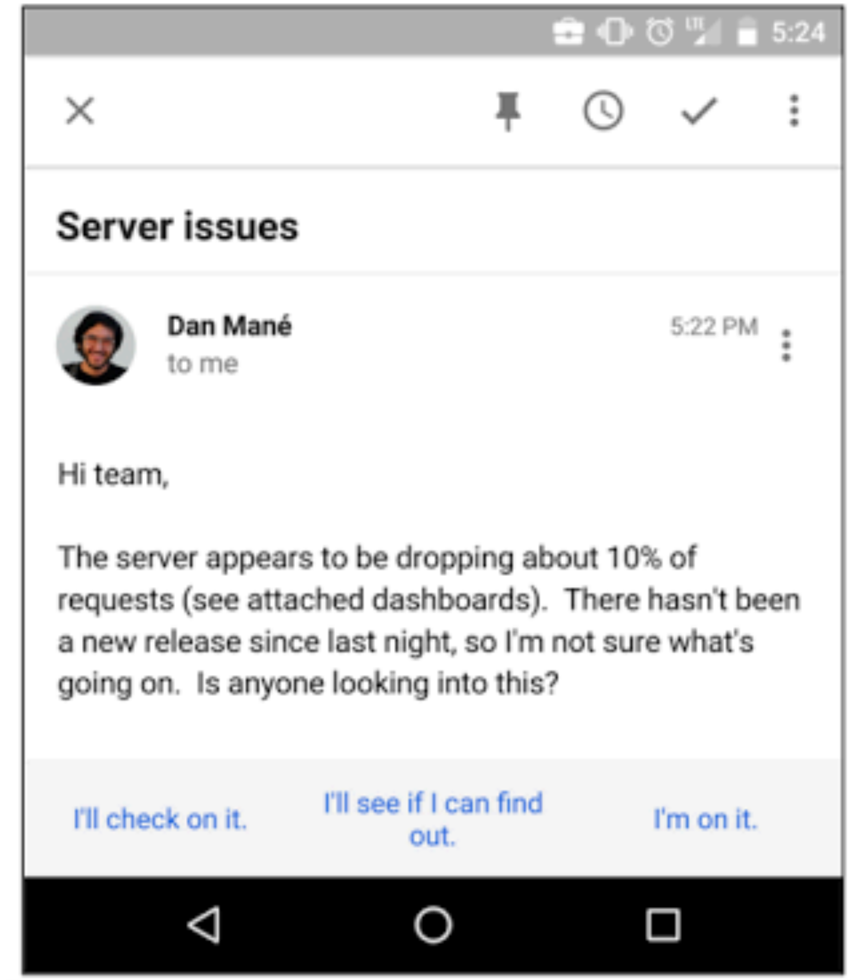
[Li et. al. 2016] [Kannan et. al. 2016] [Serban et. al. 2016]

Google Research Blog

Computer, respond to this email.

Tuesday, November 03, 2015

Posted by Greg Corrado*, Senior Research Scientist



Another bizarre feature of our early prototype was its propensity to respond with “I love you” to seemingly anything. As adorable as this sounds, it wasn’t really what we were hoping for. Some analysis revealed that the system was doing exactly what we’d trained it to do, generate likely responses -- and it turns out that responses like “Thanks", "Sounds good", and “I love you” are super common -- so the system would lean on them as a safe bet if it was unsure. Normalizing the

Neural Conversation

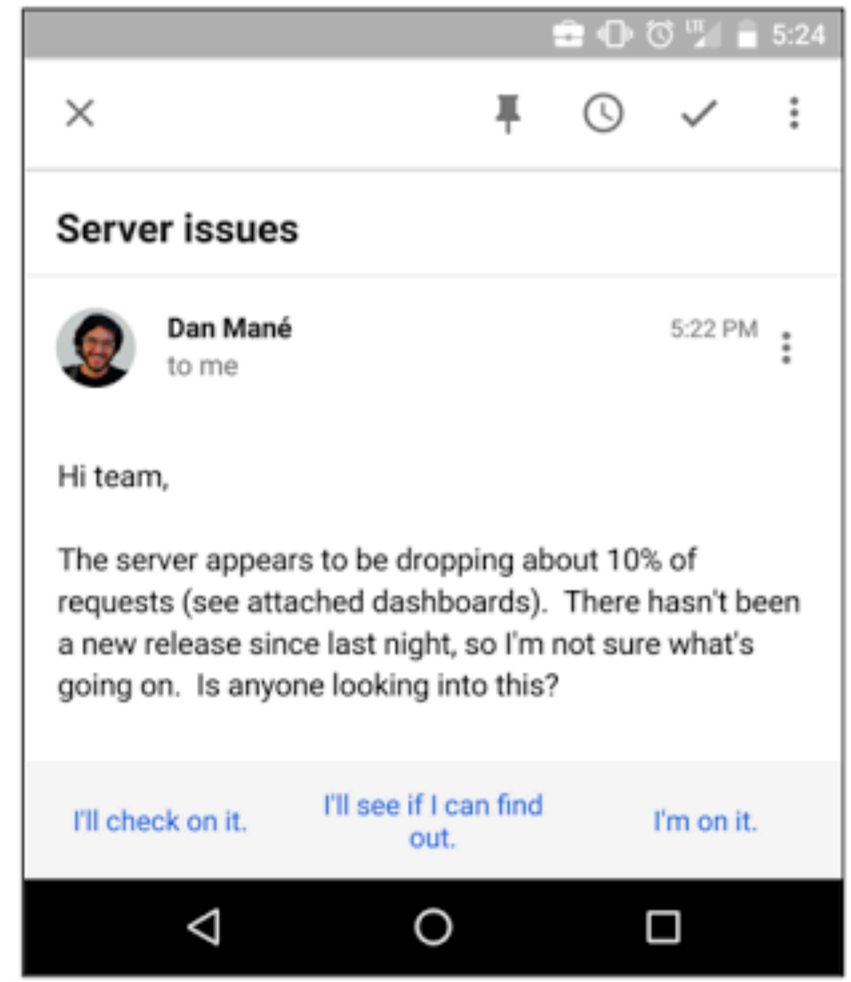
[Sordoni et. al. 2015] [Xu et. al. 2016] [Wen et. al. 2016]
[Li et. al. 2016] [Kannan et. al. 2016] [Serban et. al. 2016]

Google Research Blog

Computer, respond to this email.

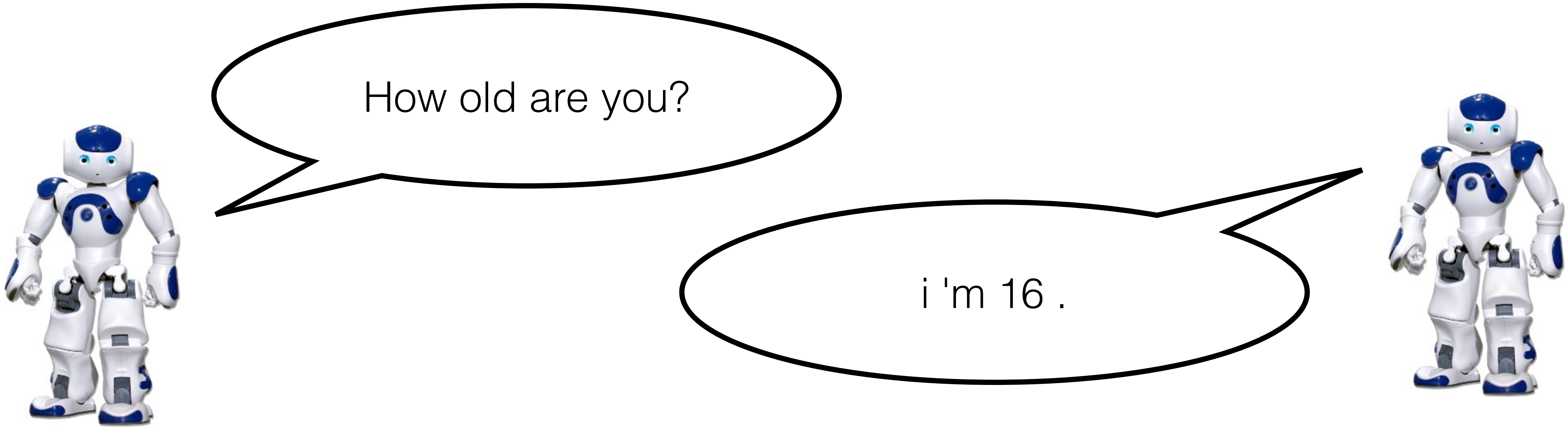
Tuesday, November 03, 2015

Posted by Greg Corrado*, Senior Research Scientist

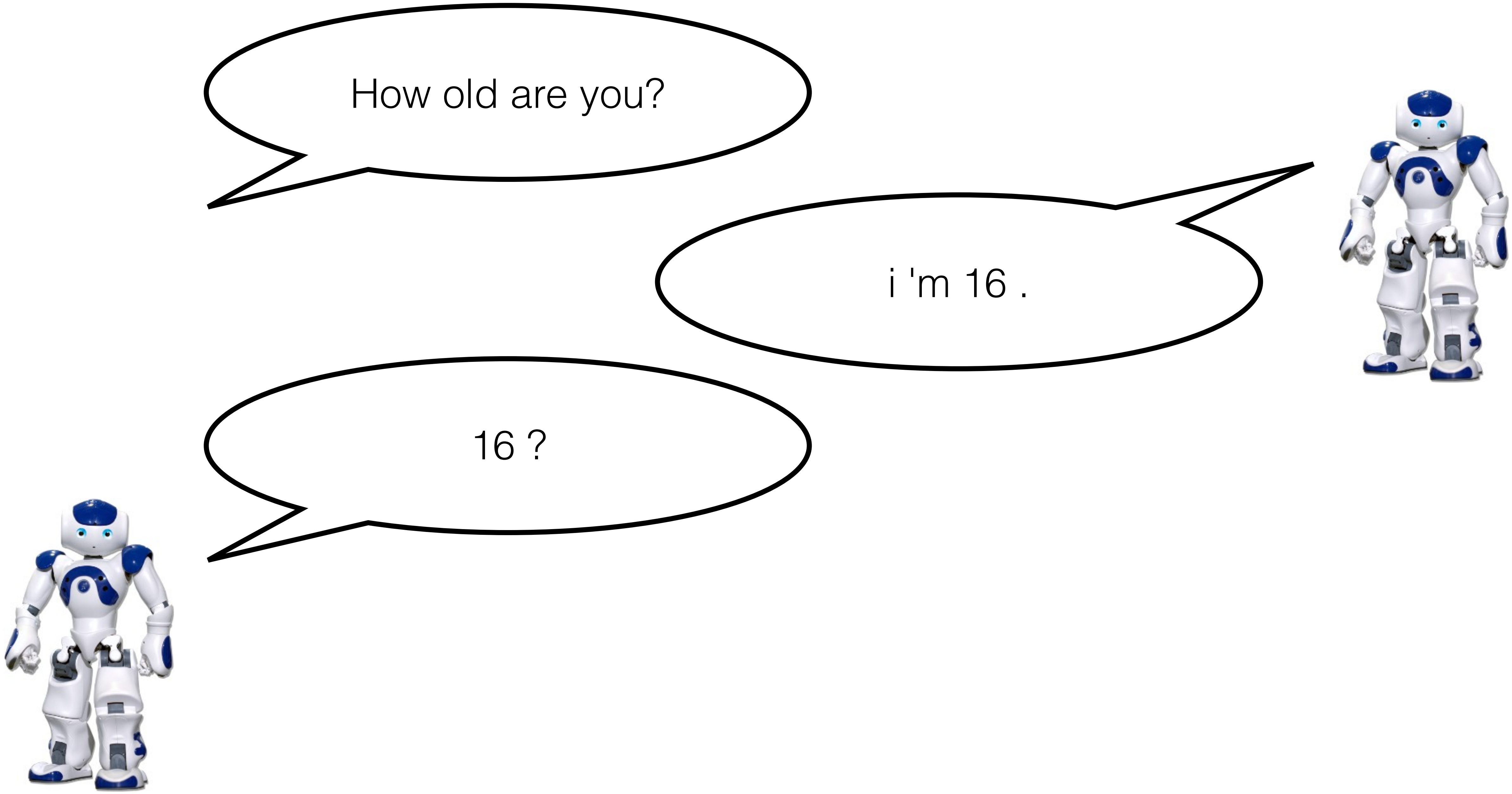


Another bizarre feature of our early prototype was its propensity to respond with “I love you” to seemingly anything. As adorable as this sounds, it wasn’t really what we were hoping for. Some analysis revealed that the system was doing exactly what we’d trained it to do, generate likely responses -- and it turns out that responses like “Thanks", "Sounds good", and “I love you” are super common -- so the system would lean on them as a safe bet if it was unsure. Normalizing the

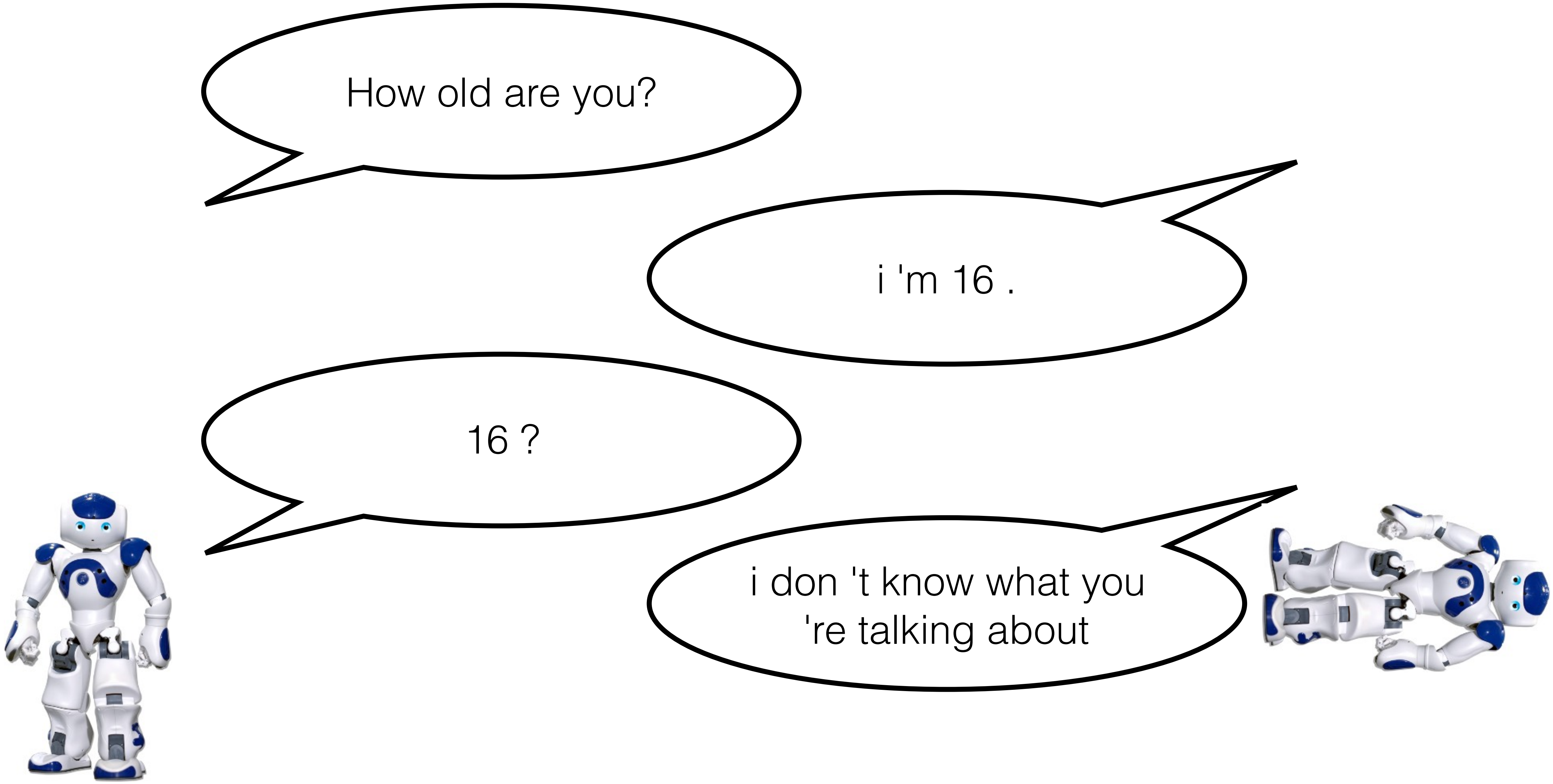
Slide Credit: Jiwei Li



Slide Credit: Jiwei Li



Slide Credit: Jiwei Li



Slide Credit: Jiwei Li

How old are you?

i 'm 16 .

16 ?

i don 't know what you 're talking about

you don 't know what you 're saying

How old are you?

i 'm 16 .

16 ?

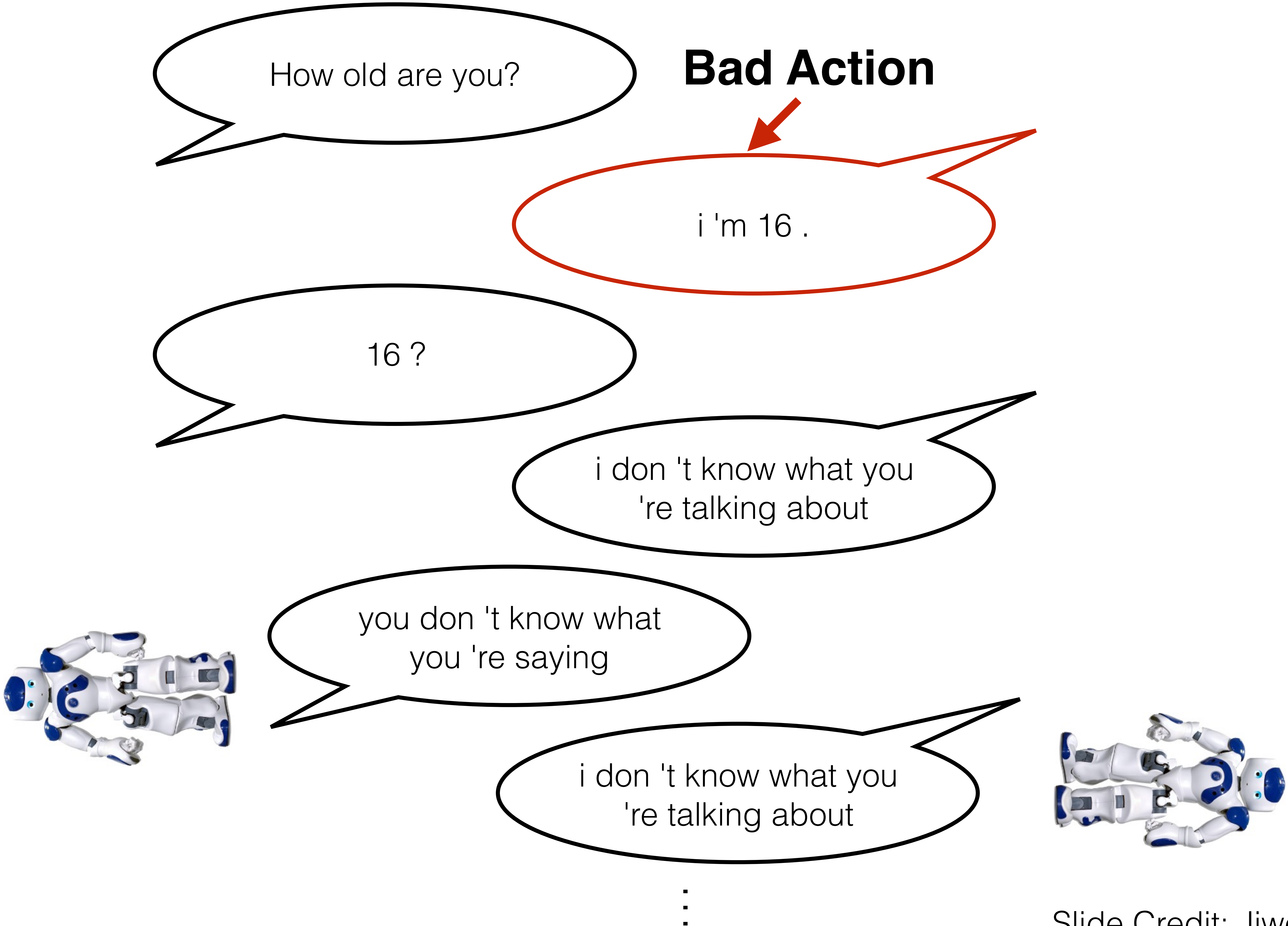
i don 't know what you 're talking about

you don 't know what you 're saying

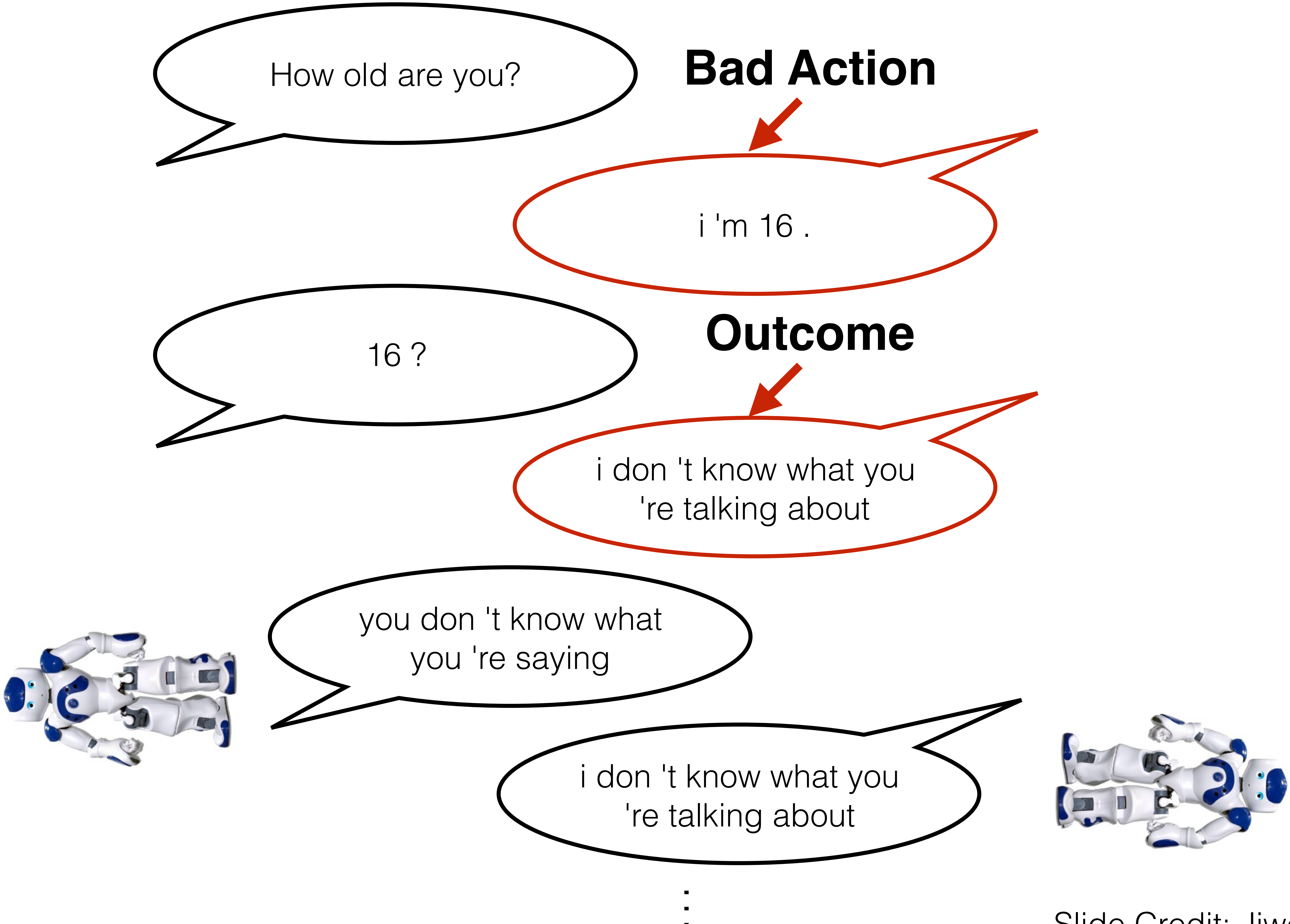
i don 't know what you 're talking about

⋮

Slide Credit: Jiwei Li



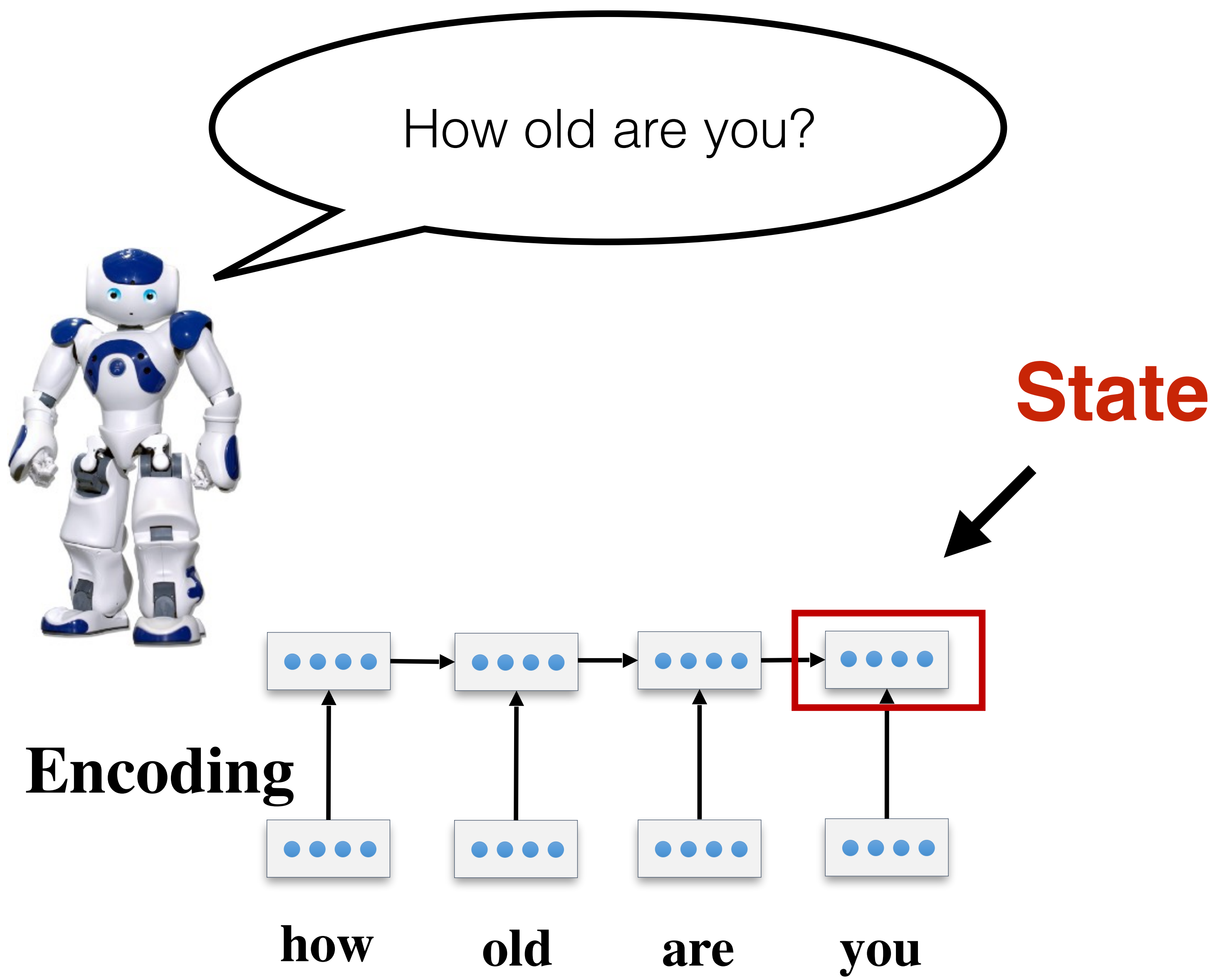
Slide Credit: Jiwei Li



Slide Credit: Jiwei Li

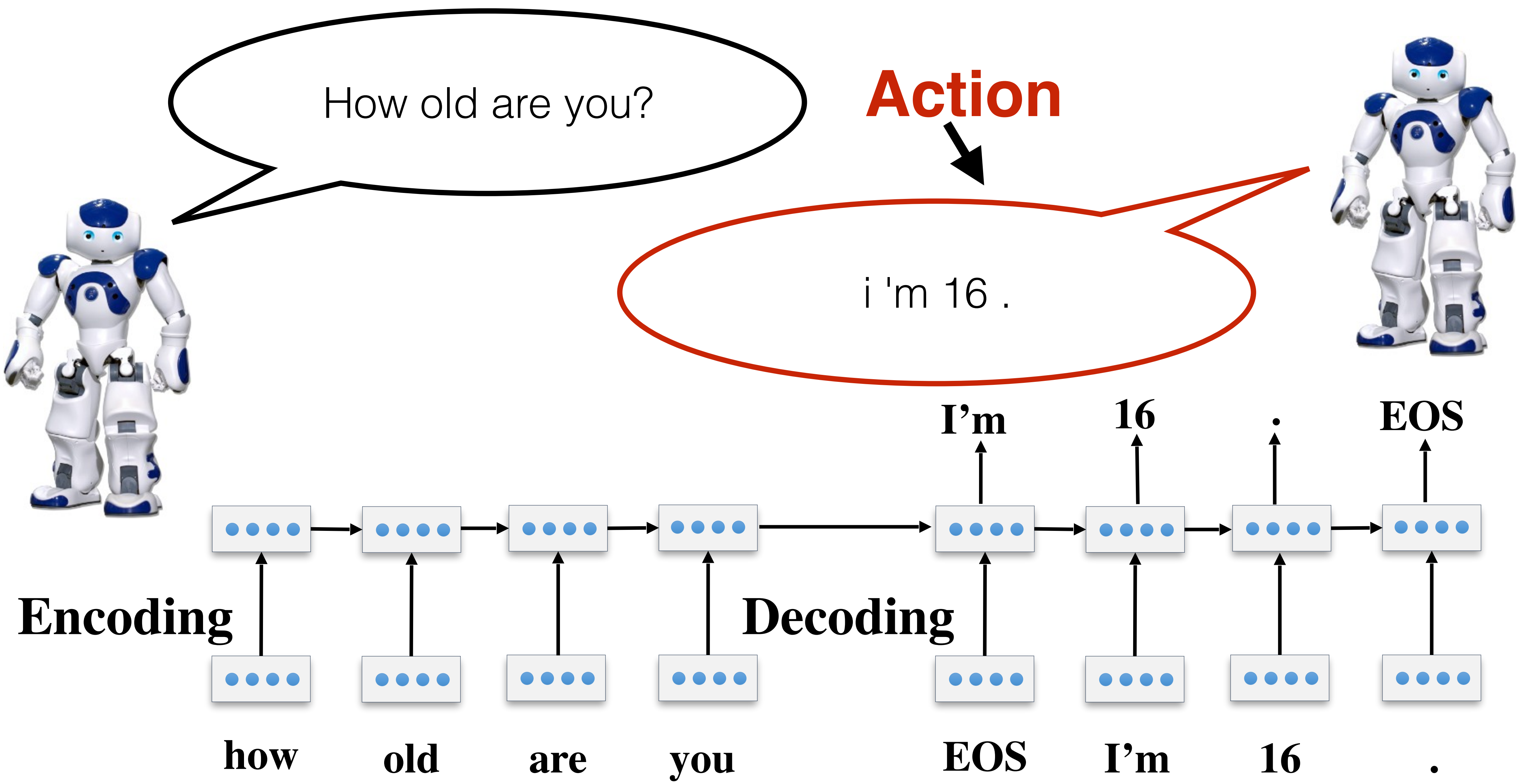
Deep Reinforcement Learning

[Li, Monroe, Ritter, Galley, Gao, Jurafsky EMNLP 2016]



Deep Reinforcement Learning

[Li, Monroe, Ritter, Galley, Gao, Jurafsky EMNLP 2016]



Learning: Policy Gradient

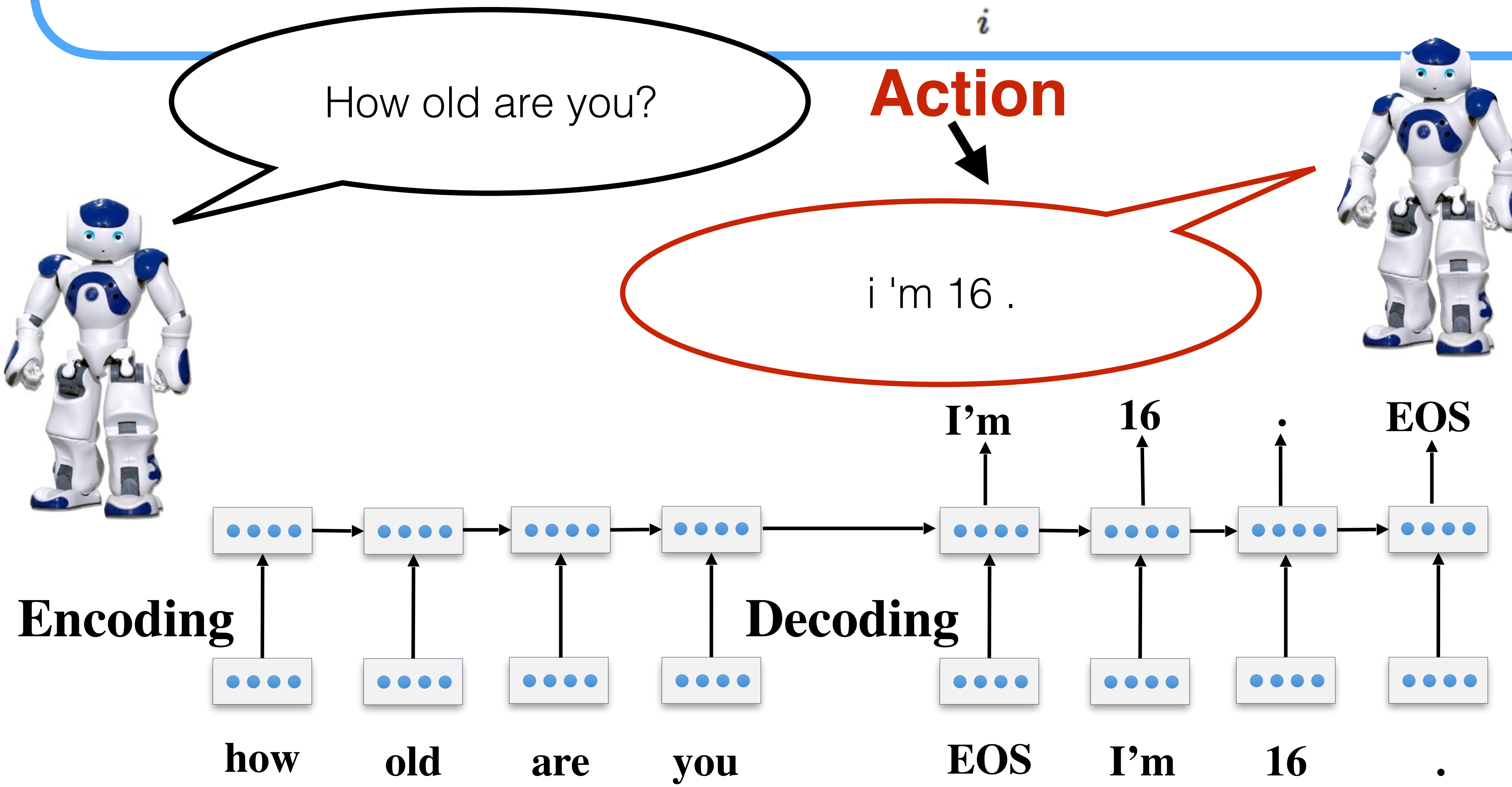
REINFORCE Algorithm (Williams,1992)

$$J(\theta) = \mathbb{E}[R(s_1, s_2, ..., s_N)]$$

$$\nabla J(\theta) = \nabla \log p(s_1, s_2, ..., s_N) R(s_1, s_2, ..., s_N)$$

What we want to learn

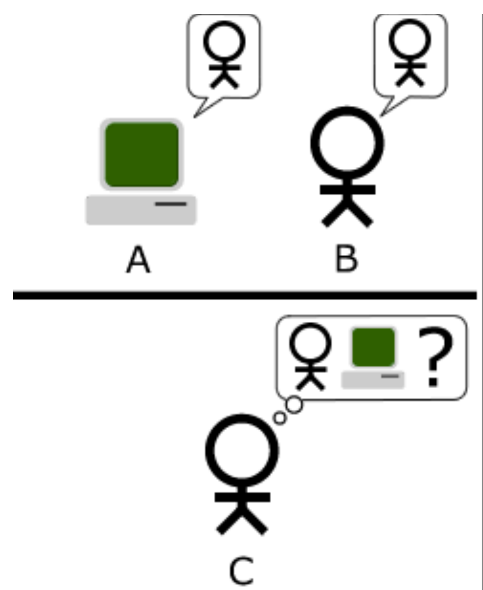
$$\nabla J(\theta) = \nabla \log \prod_i \boxed{p(s_i|s_{i-1})} R(s_1, s_2, ..., s_N)$$

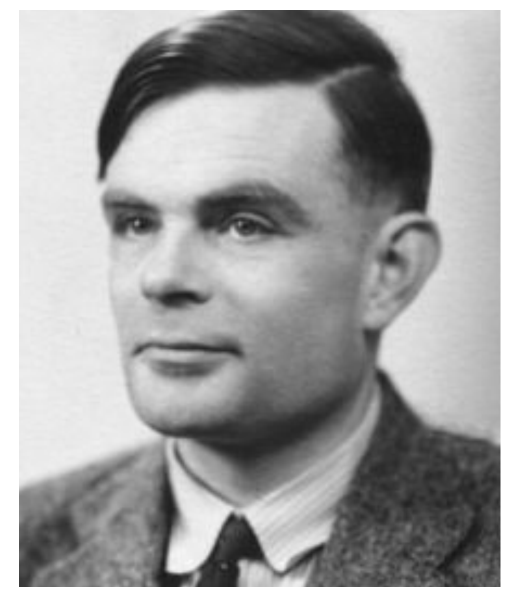


Q: Rewards?

Q: Rewards?

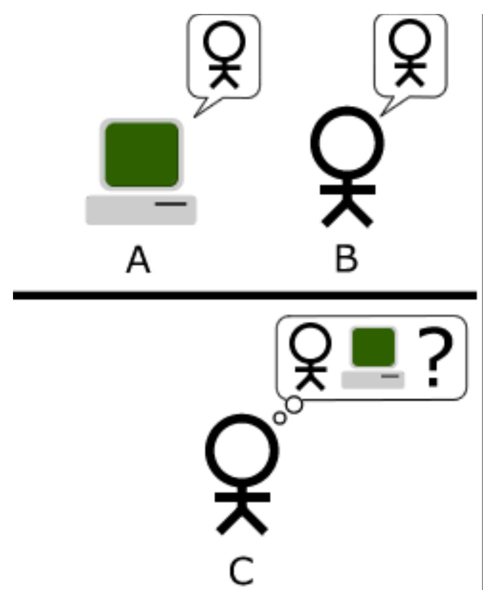
A: Turing Test

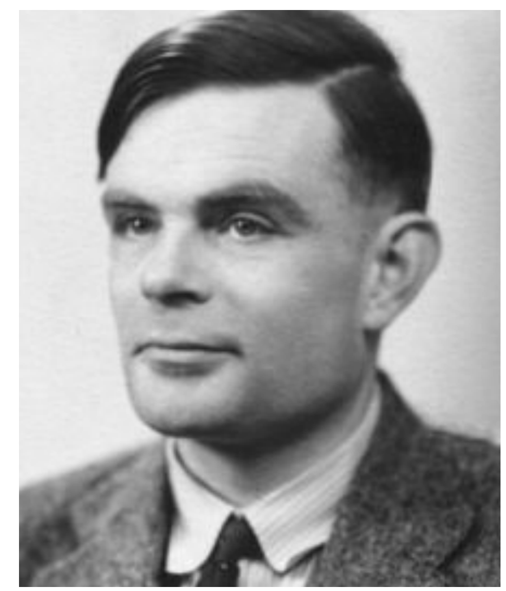




Q: Rewards?

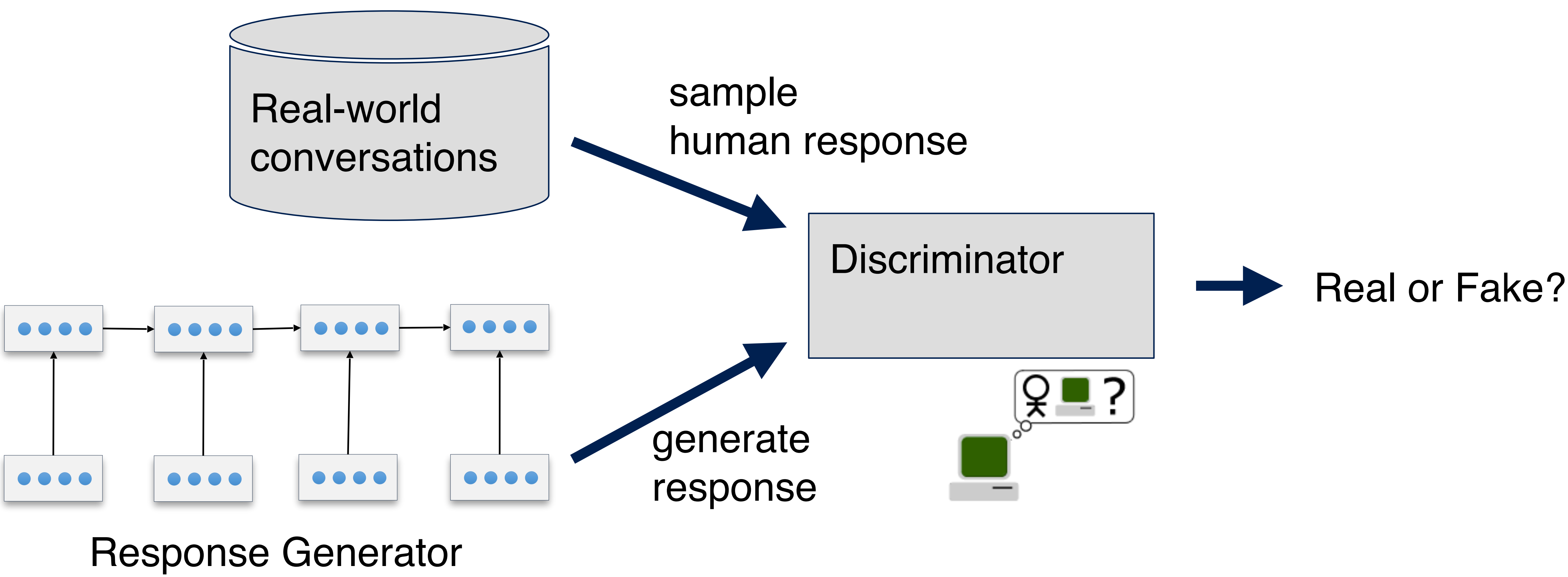
A: Turing Test





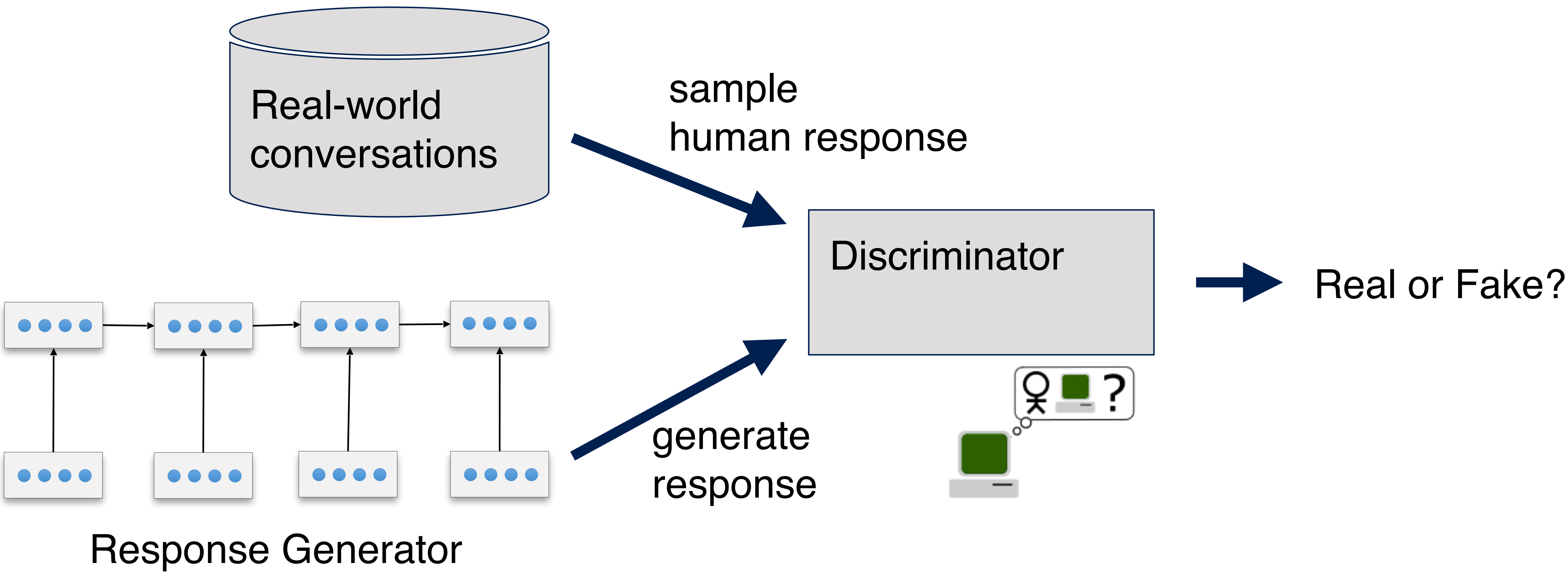
Adversarial Learning
(Goodfellow et al., 2014)

Adversarial Learning for Neural Dialogue [Li, Monroe, Shi, Jean, Ritter, Jurafsky EMNLP 2016]



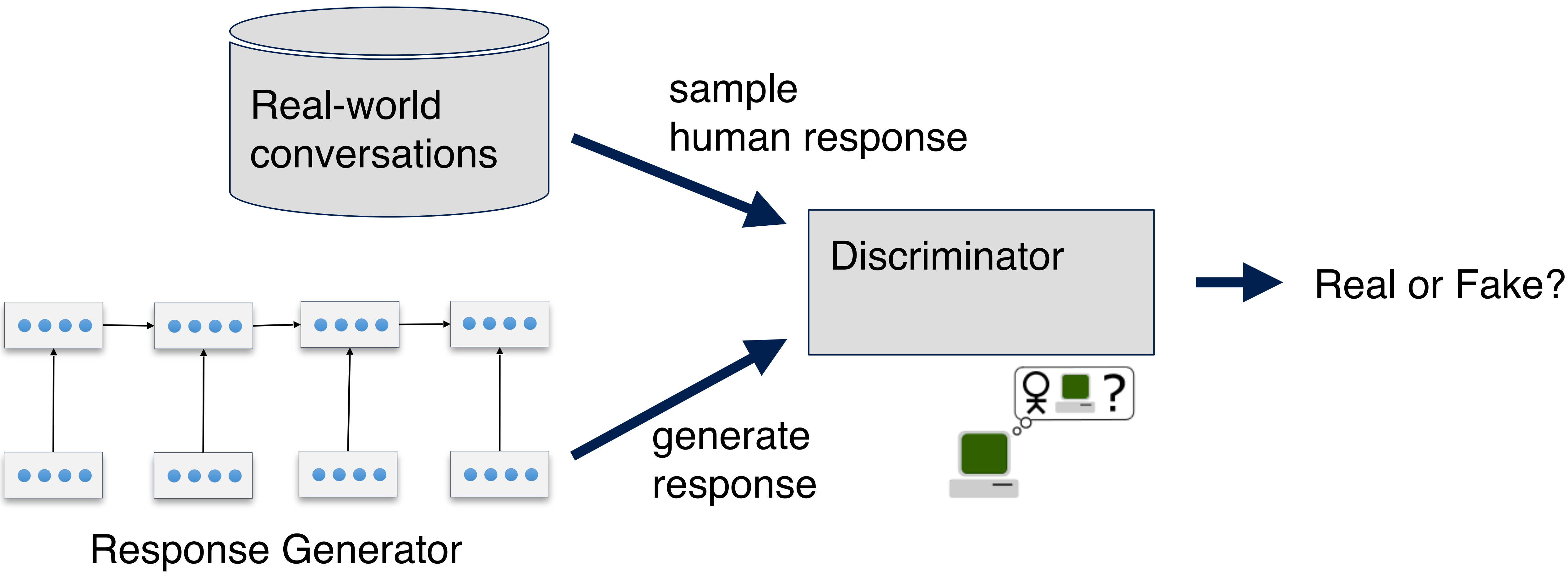
Adversarial Learning for Neural Dialogue [Li, Monroe, Shi, Jean, Ritter, Jurafsky EMNLP 2016]

(Alternate Between Training Generator and Discriminator)



Adversarial Learning for Neural Dialogue [Li, Monroe, Shi, Jean, Ritter, Jurafsky EMNLP 2016]

(Alternate Between Training Generator and Discriminator)



REINFORCE Algorithm (Williams, 1992)

Adversarial Learning Improves Response Generation

Human Evaluator:

vs vanilla generation model

Adversarial Win	Adversarial Lose	Tie
62%	18%	20%

Machine Evaluator:
[Bowman et. al. 2016]

Adversarial Success
(How often can you fool a machine)

Adversarial Learning	8.0%
Standard Seq2Seq model	4.9%

Slide Credit: Jiwei Li

Q: Why are we so good at Speech, MT (but bad at NLU)?

People ***naturally*** translate and transcribe.

Q: Large, End-to-End Datasets for NLU?

- Web-scale Conversations?
- Web-scale Structured Data?

Q: Why are we so good at Speech, MT (but bad at NLU)?

People ***naturally*** translate and transcribe.

Q: Large, End-to-End Datasets for NLU?

- Web-scale Conversations?
- Web-scale Structured Data?

Generates fluent open domain replies

Q: Why are we so good at Speech, MT (but bad at NLU)?

People ***naturally*** translate and transcribe.

Q: Large, End-to-End Datasets for NLU?

- Web-scale Conversations?
- Web-scale Structured Data?

Generates fluent open domain replies

Really Natural Language Understanding?

Q: Why are we so good at Speech, MT (but bad at NLU)?

People ***naturally*** translate and transcribe.

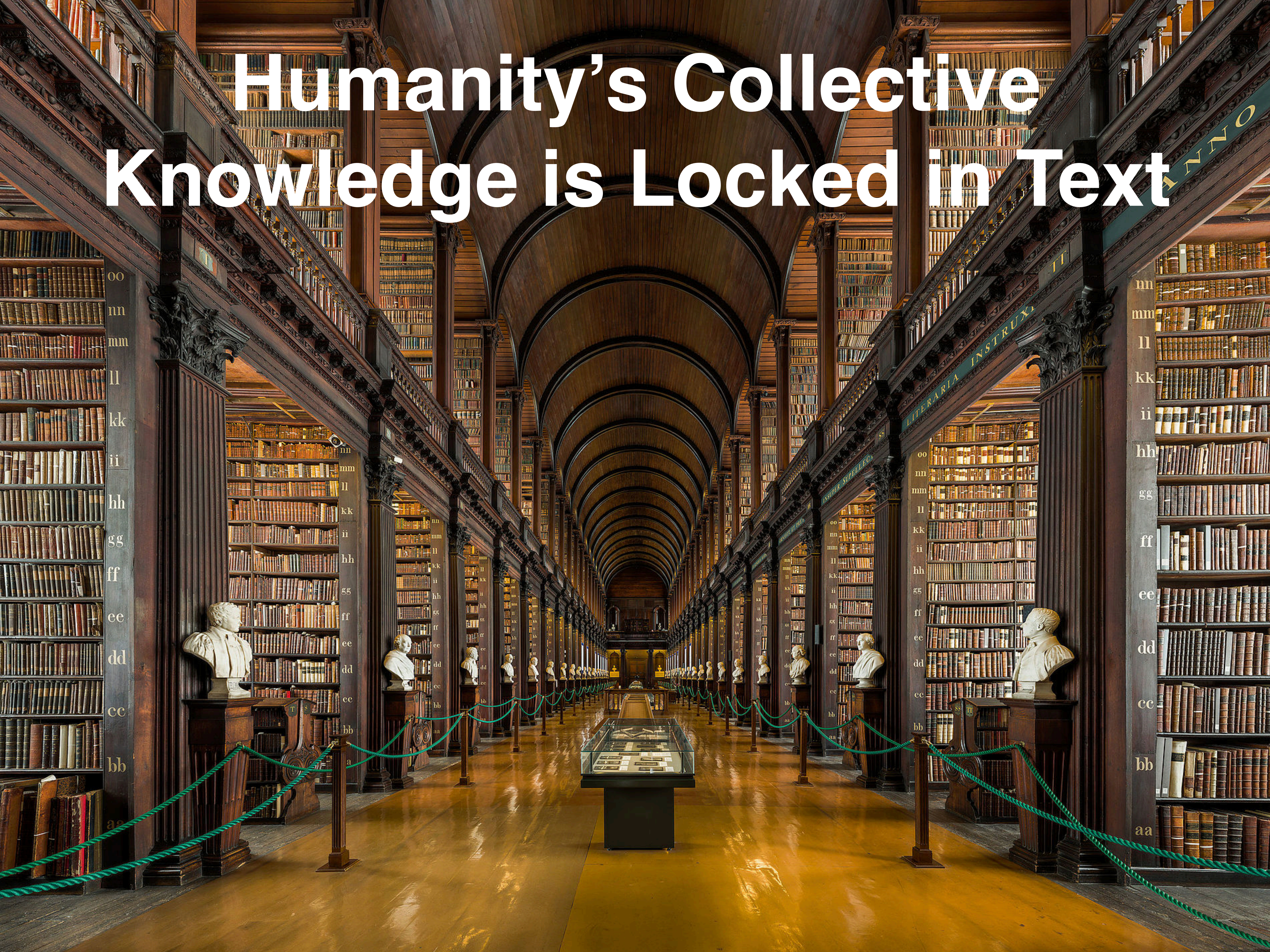
Q: Large, End-to-End Datasets for NLU?

- Web-scale Conversations?
- Web-scale Structured Data?

Generates fluent open domain replies

Really Natural Language Understanding?

Humanity’s Collective Knowledge is Locked in Text



Traditional Information Extraction

Traditional Information Extraction

1. Humans Annotate Text

Traditional Information Extraction

1. Humans Annotate Text

2. Supervised Machine Learning

$$\frac{1}{Z(w_1, \ldots, w_n, \theta)} \prod_{i=1}^{n} e^{\theta \cdot f(t_i, t_{i-1}, w_1, \ldots, w_n, i)}$$

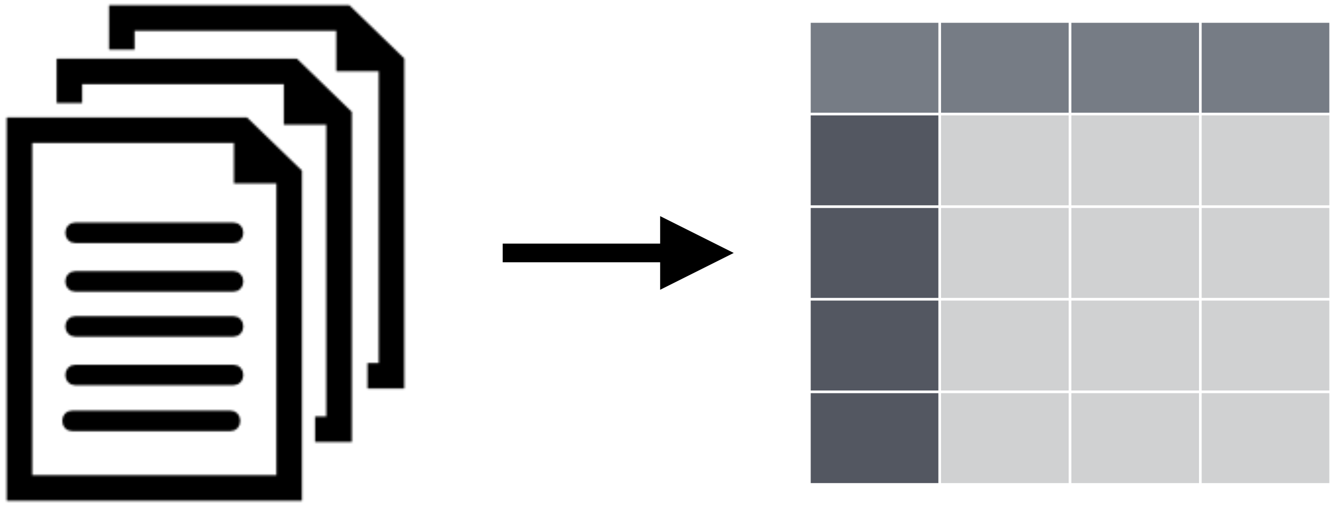
Traditional Information Extraction

1. Humans Annotate Text

2. Supervised Machine Learning

$$\frac{1}{Z(w_1,\ldots,w_n,\theta)}\prod_{i=1}^{n} e^{\theta\cdot f(t_i,t_{i-1},w_1,\ldots,w_n,i)}$$

3. Apply Models to New Documents



Traditional Information Extraction

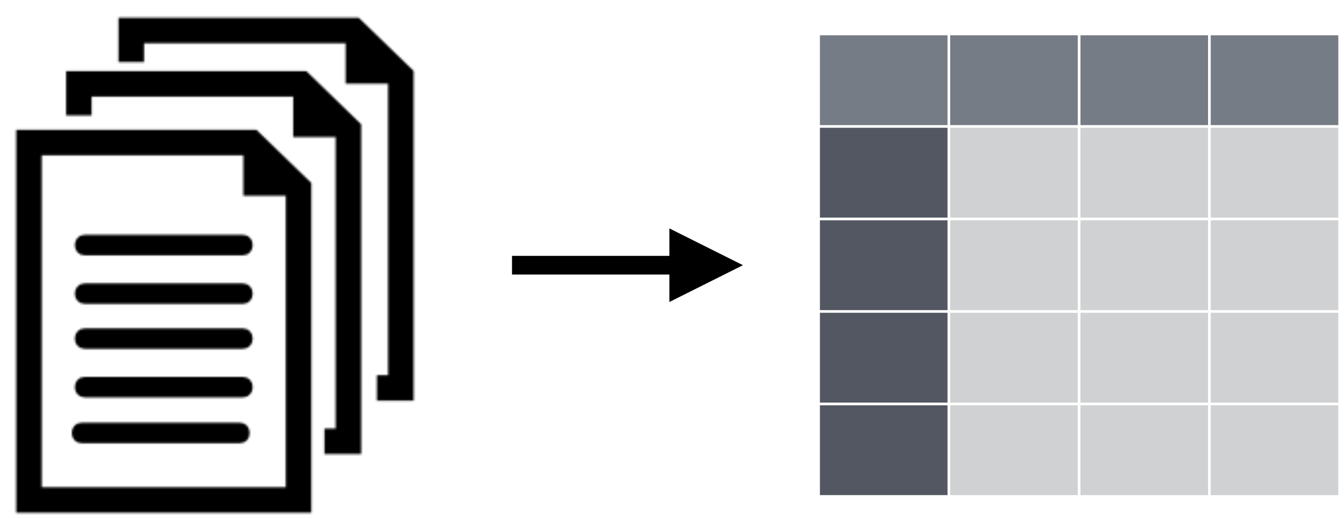
1. Humans Annotate Text

2. Supervised Machine Learning

$$\frac{1}{Z(w_1,\ldots,w_n,\theta)}\prod_{i=1}^{n} e^{\theta\cdot f(t_i,t_{i-1},w_1,\ldots,w_n,i)}$$

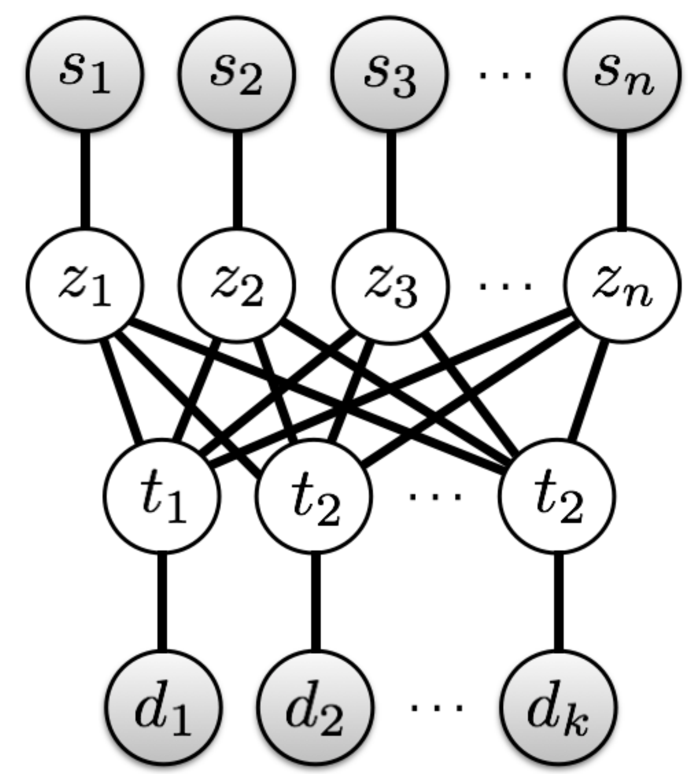
Major Bottleneck (e.g. ACE corpus contains ~400 newswire documents)

3. Apply Models to New Documents



Learning from **Distant Supervision**

[Mintz et. al. 2009]



1) Named Entity Recognition

Challenge: highly ambiguous labels

[Ritter, et. al. EMNLP 2011]

2) Relation Extraction

Challenge: missing data

[Ritter, et. al. TACL 2013]

3) Time Normalization

Challenge: diversity in noisy text

[Tabassum, Ritter, Xu, EMNLP 2016]

4) Event Extraction

Challenge: lack of negative examples

[Ritter, et. al. WWW 2015]

[Konovalov, et. al. WWW 2017]

$$O(\theta) = \underbrace{\sum_{i}^{N} \log p_\theta(y_i|x_i)}_{\text{Log Likelihood}} - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Learning from **Distant Supervision**

[Mintz et. al. 2009]

1) Named Entity Recognition

Challenge: highly ambiguous labels

[Ritter, et. al. EMNLP 2011]

2) Relation Extraction

Challenge: missing data

[Ritter, et. al. TACL 2013]

3) Time Normalization

Challenge: diversity in noisy text

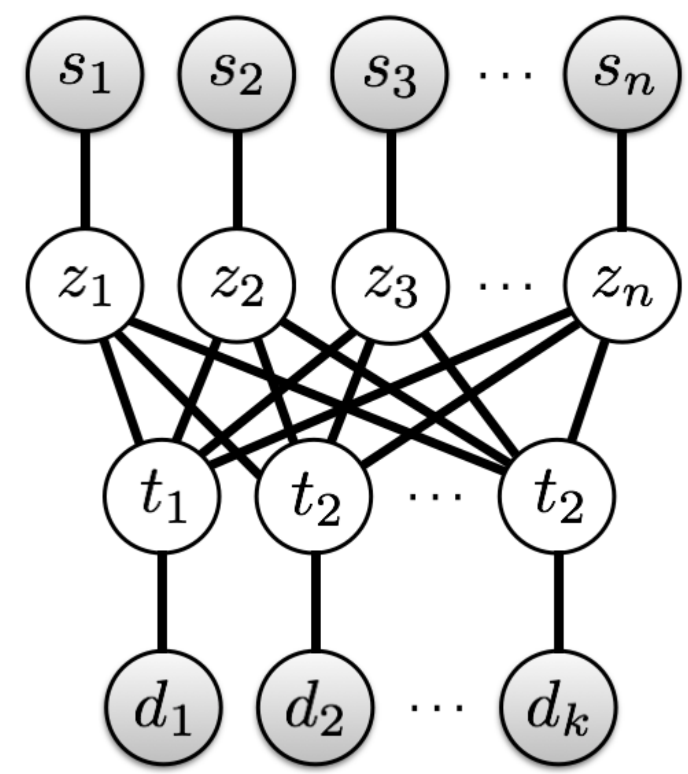
[Tabassum, Ritter, Xu, EMNLP 2016]

4) Event Extraction

Challenge: lack of negative examples

[Ritter, et. al. WWW 2015]

[Konovalov, et. al. WWW 2017]



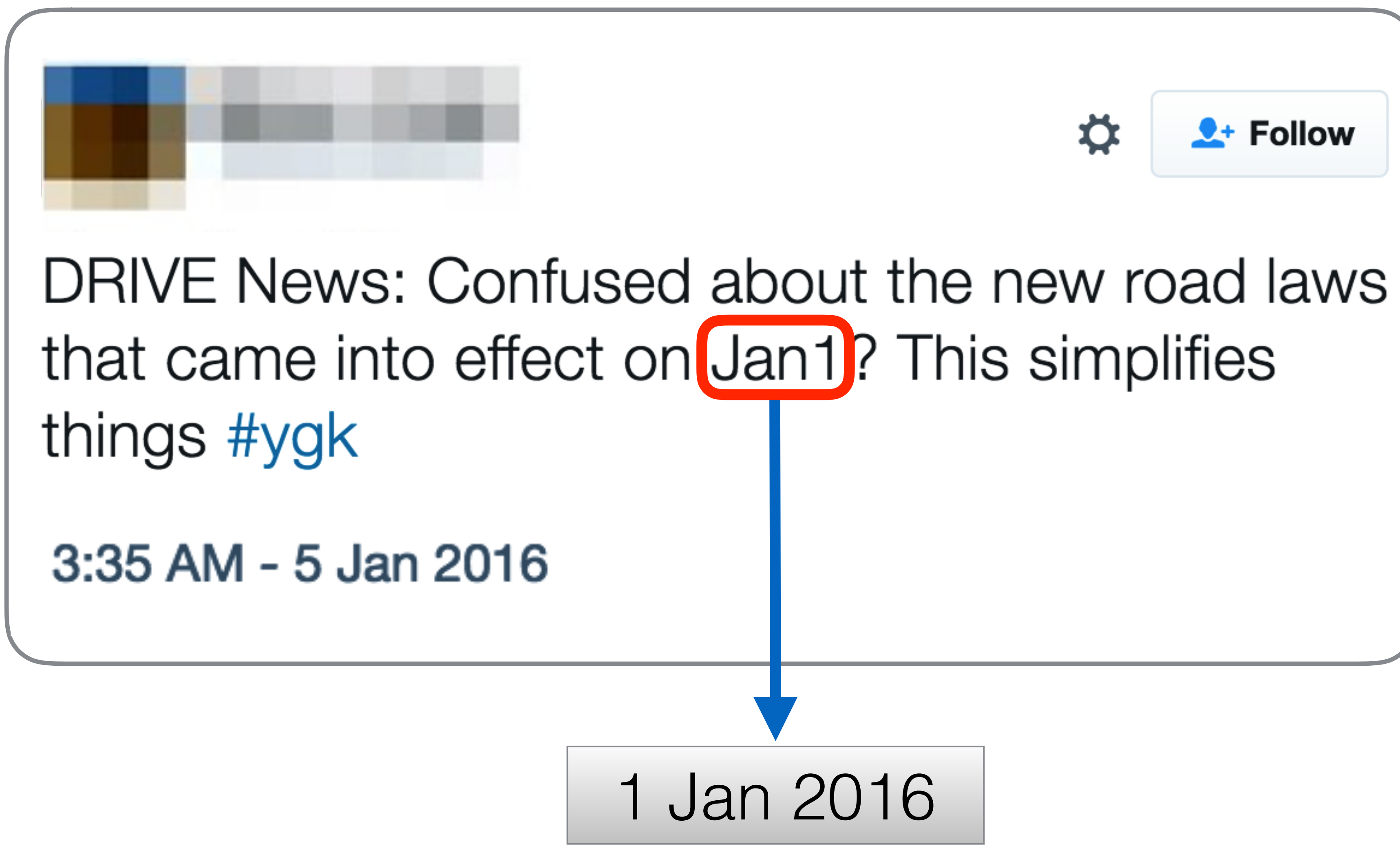
$$O(\theta) = \underbrace{\sum_{i}^{N} \log p_\theta(y_i|x_i)}_{\text{Log Likelihood}} - \underbrace{\lambda^U D(\tilde{p}||\hat{p}_\theta^{\text{unlabeled}})}_{\text{Label regularization}}$$

Time Normalization

[Tabassum, Ritter, Xu EMNLP 2016]

State-of-the-art time resolvers

- TempEX
- HeidelTime
- SUTime
- UWTime



Time Normalization

[Tabassum, Ritter, Xu EMNLP 2016]

Distant Supervision
(no human labels or rules!)

State-of-the-art time resolvers

- TempEX
- HeidelTime
- SUTime
- UWTime

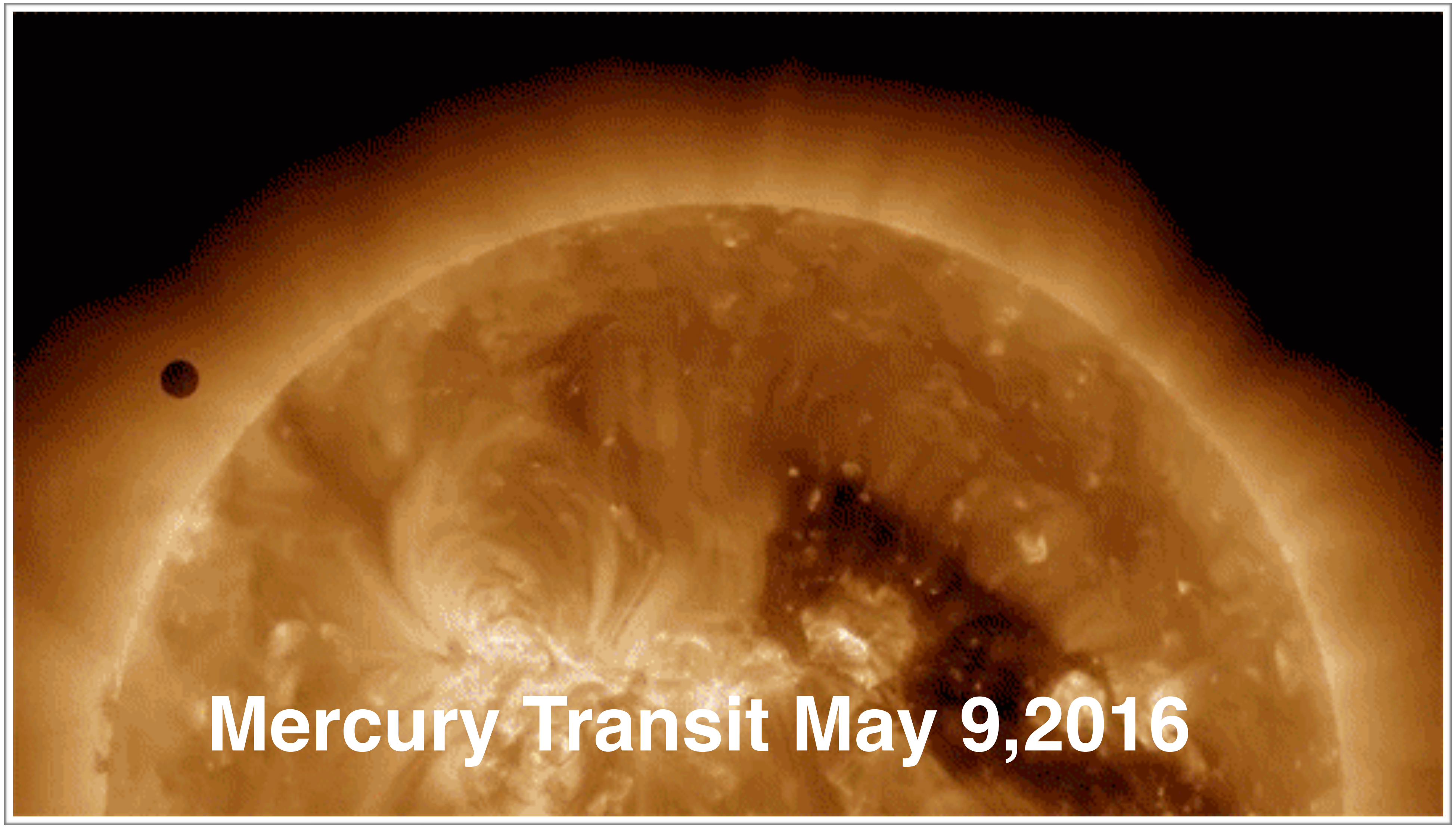
Follow

DRIVE News: Confused about the new road laws that came into effect on Jan1? This simplifies things #ygk

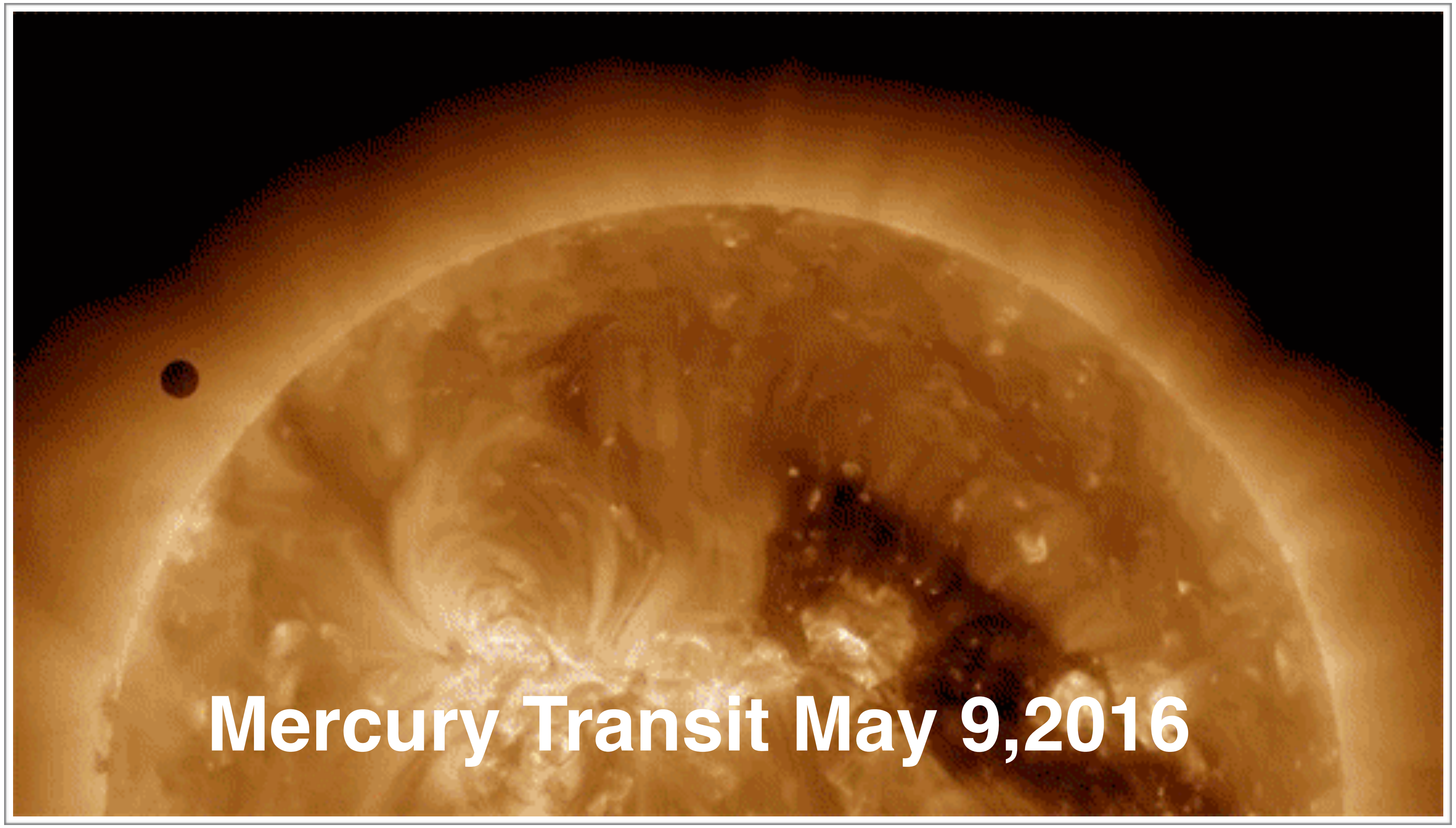
3:35 AM - 5 Jan 2016

1 Jan 2016

Distant Supervision Assumption

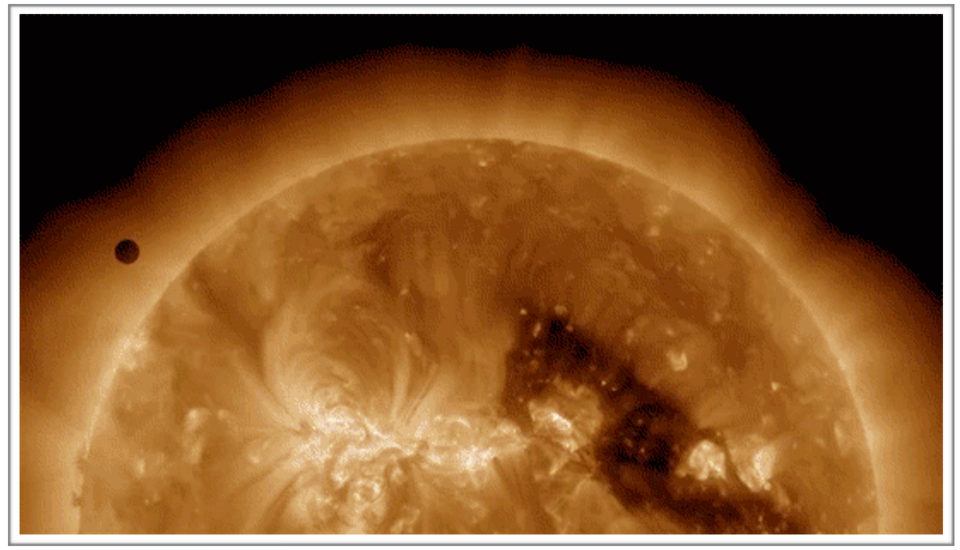


Distant Supervision Assumption



Distant Supervision Assumption

Mercury Transit
May 9,2016



8 May

Mercury will make a rare transit across the sun tmrw morning (Mon). If you're able to catch it, don't miss out -- and use a solar filter!

10:28 PM - 8 May 2016

Mercury Transit 2morrow starting at 6:00 AM Mercury will pass in front of Sun @14News @14FirstAlert #mercurytransit

7:30 PM - 8 May 2016

9 May

Paul from Creators Hand Photography captured a shot of today's Mercury transit, along with a larger sunspot that... fb.me/7jaxf4rfC

i didn't get to see mercury transit today because of this horrible weather 😤

3:54 PM - 9 May 2016

10 May

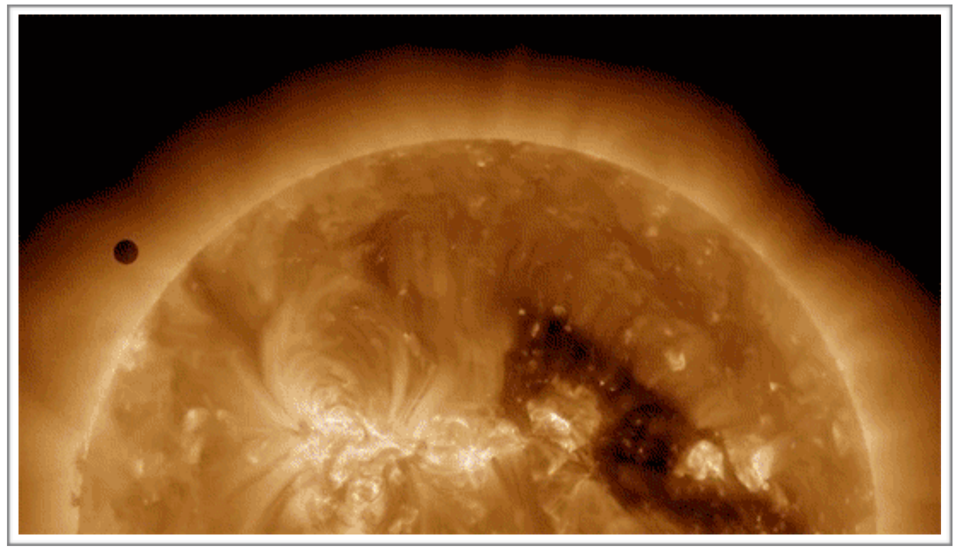
I watched this event yesterday by a small telescope with all the precautions, but this transit of Mercury is great!

Mercury passes between Earth and the sun only about 13 times a century.
It was yesterday! May 9th #lagalaxiaensmira

3:17 PM - 9 May 2016

Distant Supervision Assumption

Mercury Transit
May 9,2016



8 May

Mercury will make a rare transit across the sun tmrw morning (Mon). If you're able to catch it, don't miss out -- and use a solar filter!

10:28 PM - 8 May 2016

Mercury Transit 2morrow starting at 6:00 AM Mercury will pass in front of Sun @14News @14FirstAlert #mercurytransit

7:30 PM - 8 May 2016

9 May

Paul from Creators Hand Photography captured a shot of today's Mercury transit, along with a larger sunspot that... fb.me/7jaxf4rfC

i didn't get to see mercury transit today because of this horrible weather 😤

3:54 PM - 9 May 2016

10 May

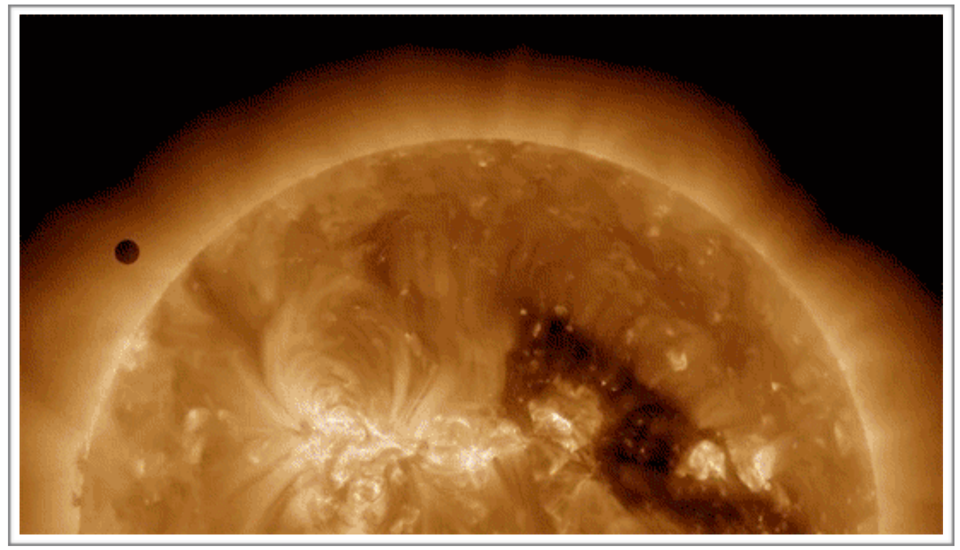
I watched this event yesterday by a small telescope with all the precautions, but this transit of Mercury is great!

Mercury passes between Earth and the sun only about 13 times a century.
It was yesterday! May 9th #lagalaxiaensmira

3:17 PM - 9 May 2016

Distant Supervision Assumption

Mercury Transit
May 9,2016



8 May

Mercury will make a rare transit across the sun tmrw morning (Mon). If you're able to catch it, don't miss out -- and use a solar filter!

10:28 PM - 8 May 2016

Mercury Transit 2morrow starting at 6:00 AM Mercury will pass in front of Sun @14News @14FirstAlert #mercurytransit

7:30 PM - 8 May 2016

9 May

Paul from Creators Hand Photography captured a shot of today's Mercury transit, along with a larger sunspot that... fb.me/7jaxf4rfC

i didn't get to see mercury transit today because of this horrible weather 😤

3:54 PM - 9 May 2016

10 May

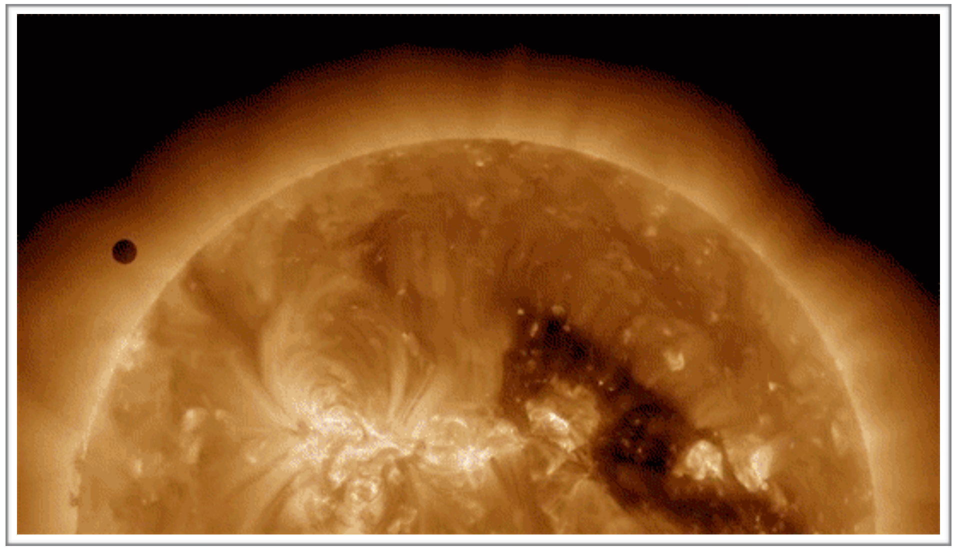
I watched this event yesterday by a small telescope with all the precautions, but this transit of Mercury is great!

Mercury passes between Earth and the sun only about 13 times a century.
It was yesterday May 9th #lagalaxiaensmira

3:17 PM - 9 May 2016

Distant Supervision Assumption

Mercury Transit
May 9,2016



8 May

Mercury will make a rare transit across the sun tmrw morning $\boxed{\text{Mon}}$. If you're able to catch it, don't miss out -- and use a solar filter!

10:28 PM - 8 May 2016

Mercury Transit $\boxed{\text{2morrow}}$ starting at 6:00 AM Mercury will pass in front of Sun @14News @14FirstAlert #mercurytransit

7:30 PM - 8 May 2016

9 May

Paul from Creators Hand Photography captured a shot of $\boxed{\text{today's}}$ Mercury transit, along with a larger sunspot that... fb.me/7jaxf4rfC

i didn't get to see mercury transit $\boxed{\text{today}}$ because of this horrible weather 😤

3:54 PM - 9 May 2016

10 May

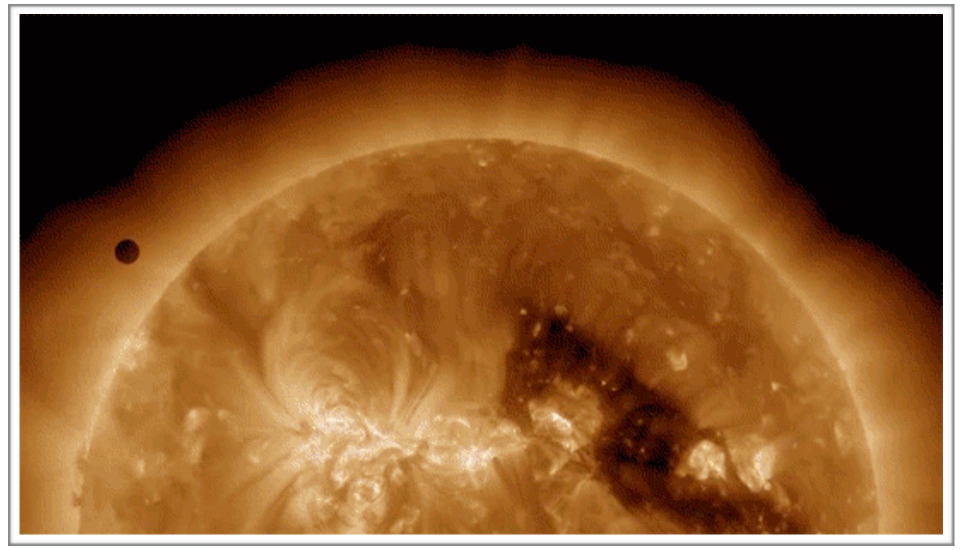
I watched this event $\boxed{\text{yesterday}}$ by a small telescope with all the precautions, but this transit of Mercury is great!

Mercury passes between Earth and the sun only about 13 times a century.
It was $\boxed{\text{yesterday}}$ $\boxed{\text{May 9th}}$ #lagalaxiaensmira

3:17 PM - 9 May 2016

Distant Supervision Assumption

Mercury Transit
May 9,2016



8 May

9 May

10 May

Mercury will make a rare transit across the sun tmrw morning (Mon). If you're able to catch it, don't miss out -- and use a solar filter!

Mercury Transit 2morrow starting at 6:00 AM Mercury will pass in front of Sun @14News @14FirstAlert #mercurytransit

7:30 PM - 8 May 2016

Paul from Creators Hand Photography captured a shot of today's Mercury transit, along with a larger sunspot that... fb.me/7jaxf4rfC

i didn't get to see mercury transit today because of this horrible weather 😤

3:54 PM - 9 May 2016

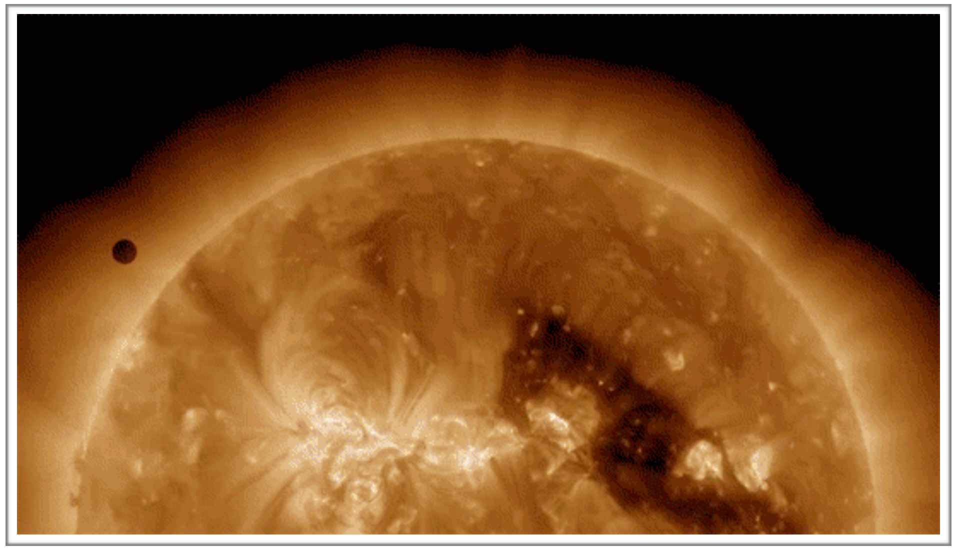
I watched this event yesterday by a small telescope with all the precautions, but this transit of Mercury is great!

Mercury passes between Earth and the sun only about 13 times a century.
It was yesterday May 9th #lagalaxiaensmira

3:17 PM - 9 May 2016

Distant Supervision Assumption

Mercury Transit
May 9,2016



8 May

Mercury will make a rare transit across the sun tmrw morning (Mon). If you're able to catch it, don't miss out -- and use a solar filter!

Mercury Transit 2morrow starting at 6:00 AM Mercury will pass in front of Sun @14News @14FirstAlert #mercurytransit

7:30 PM - 8 May 2016

9 May

Paul from Creators Hand Photography captured a shot of today's Mercury transit, along with a larger sunspot that... fb.me/7jaxf4rfC

i didn't get to see mercury transit today because of this horrible weather 😤

3:54 PM - 9 May 2016

10 May

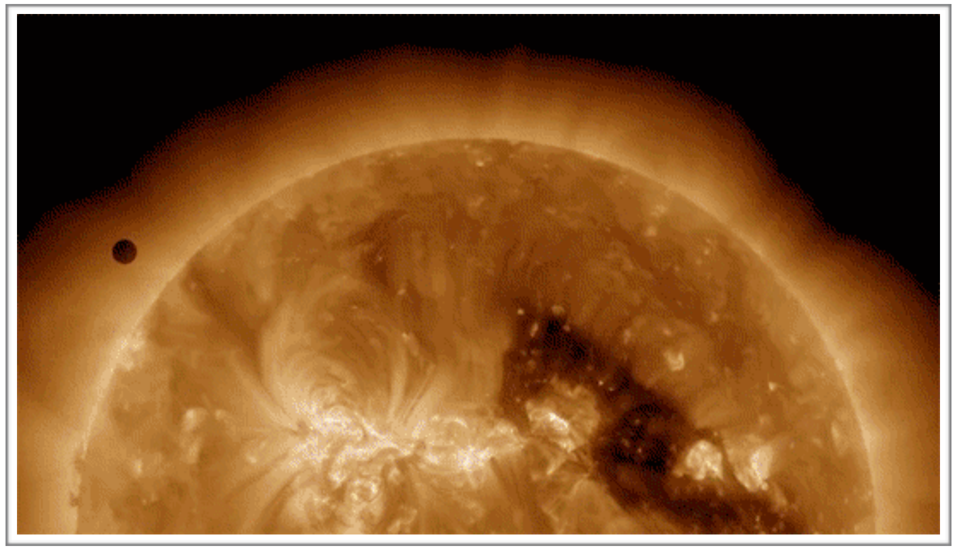
I watched this event yesterday by a small telescope with all the precautions, but this transit of Mercury is great!

Mercury passes between Earth and the sun only about 13 times a century.
It was yesterday May 9th #lagalaxiaensmira

3:17 PM - 9 May 2016

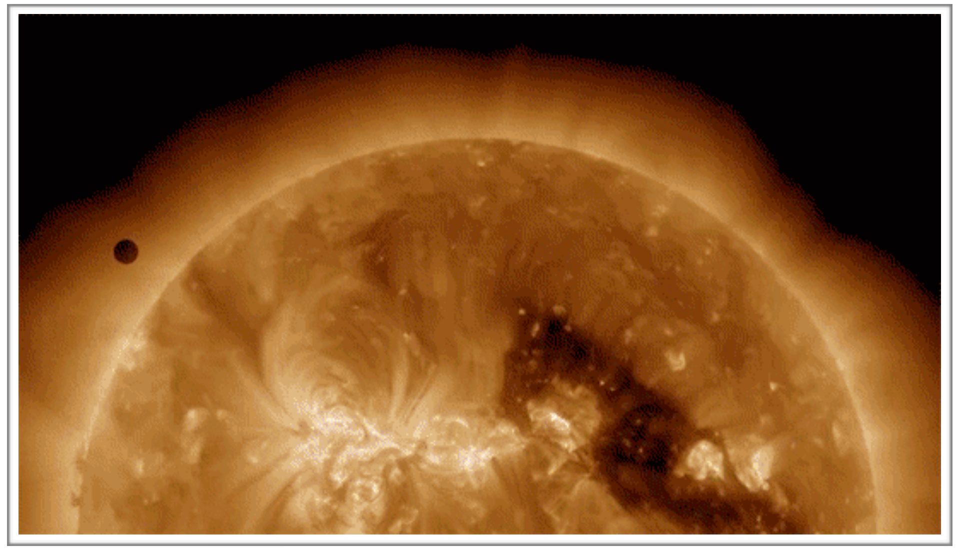
Distant Supervision Assumption

Mercury Transit
May 9,2016



Distant Supervision Assumption

Mercury Transit
May 9,2016



Distant Supervision Assumption

Tweets posted near an event that mention a key entity are likely to **contain time expressions** referring to the event's date.

Temporal Normalizer

[$\frac{1}{2}$ Million Tweets]

[NORMALIZER]

$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

Temporal Normalizer

[Event Database]

LABEL: May 9 2016

[$\frac{1}{2}$ Million Tweets]

[NORMALIZER]

$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

Temporal Normalizer

[Event Database]

LABEL: May 9 2016

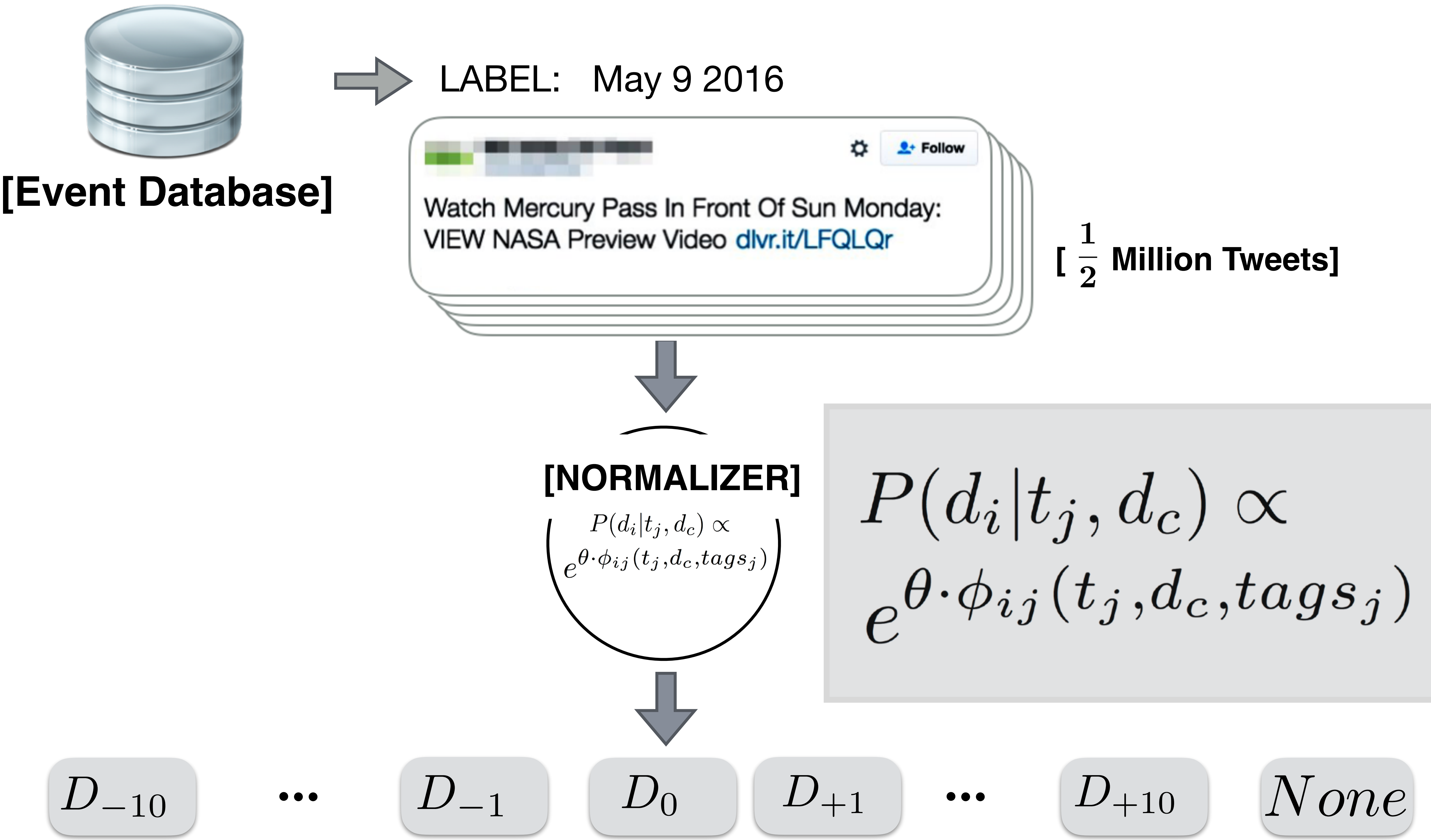
[$\frac{1}{2}$ Million Tweets]

[NORMALIZER]

$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

Temporal Normalizer



Temporal Normalizer

[Event Database]

LABEL: May 9 2016

Follow

Watch Mercury Pass In Front Of Sun Monday: VIEW NASA Preview Video dlvr.it/LFQLQr

[$\frac{1}{2}$ Million Tweets]

Creation Date: 11/1

EMNLP starts 4m tmrw!

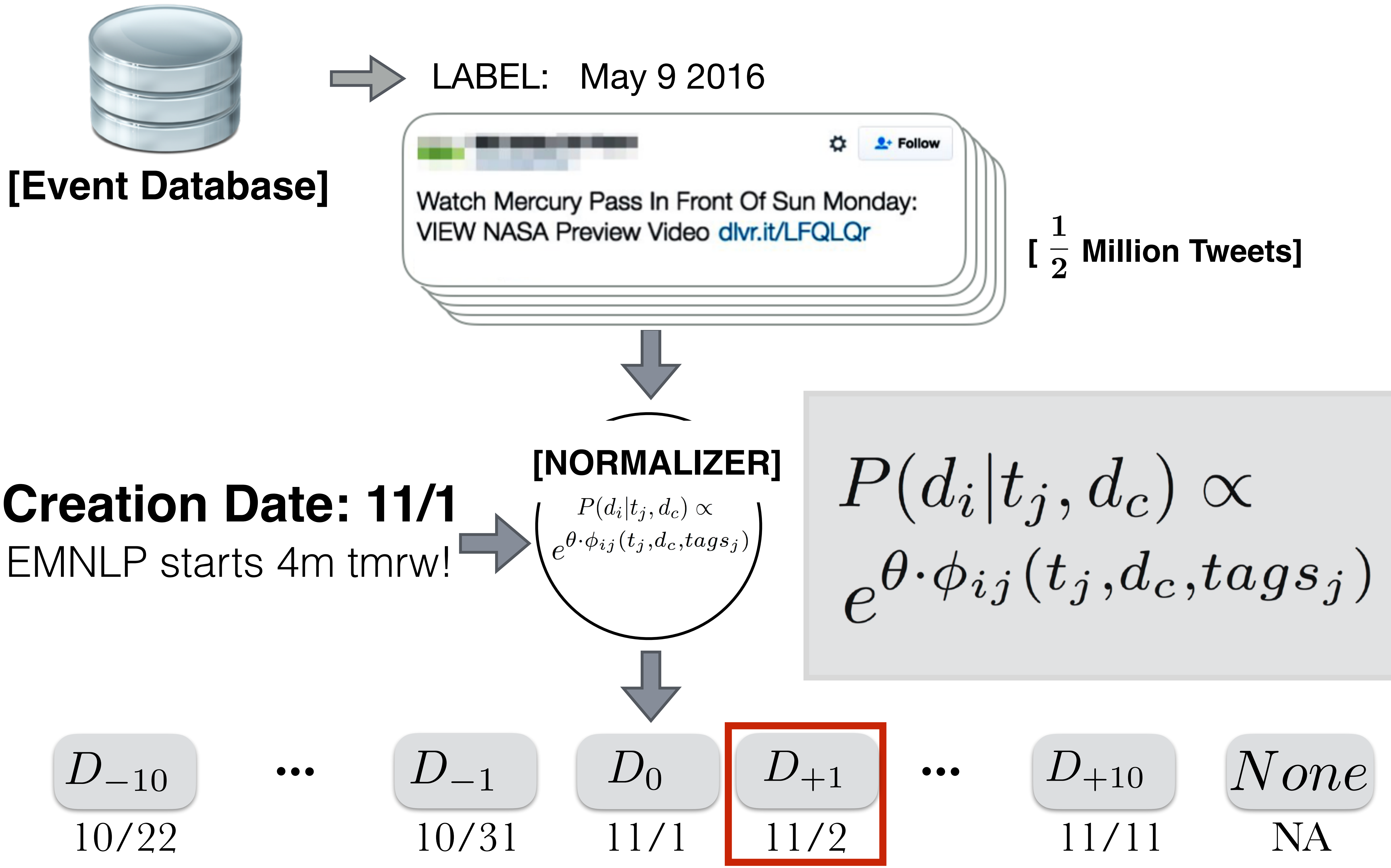
[NORMALIZER]

$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

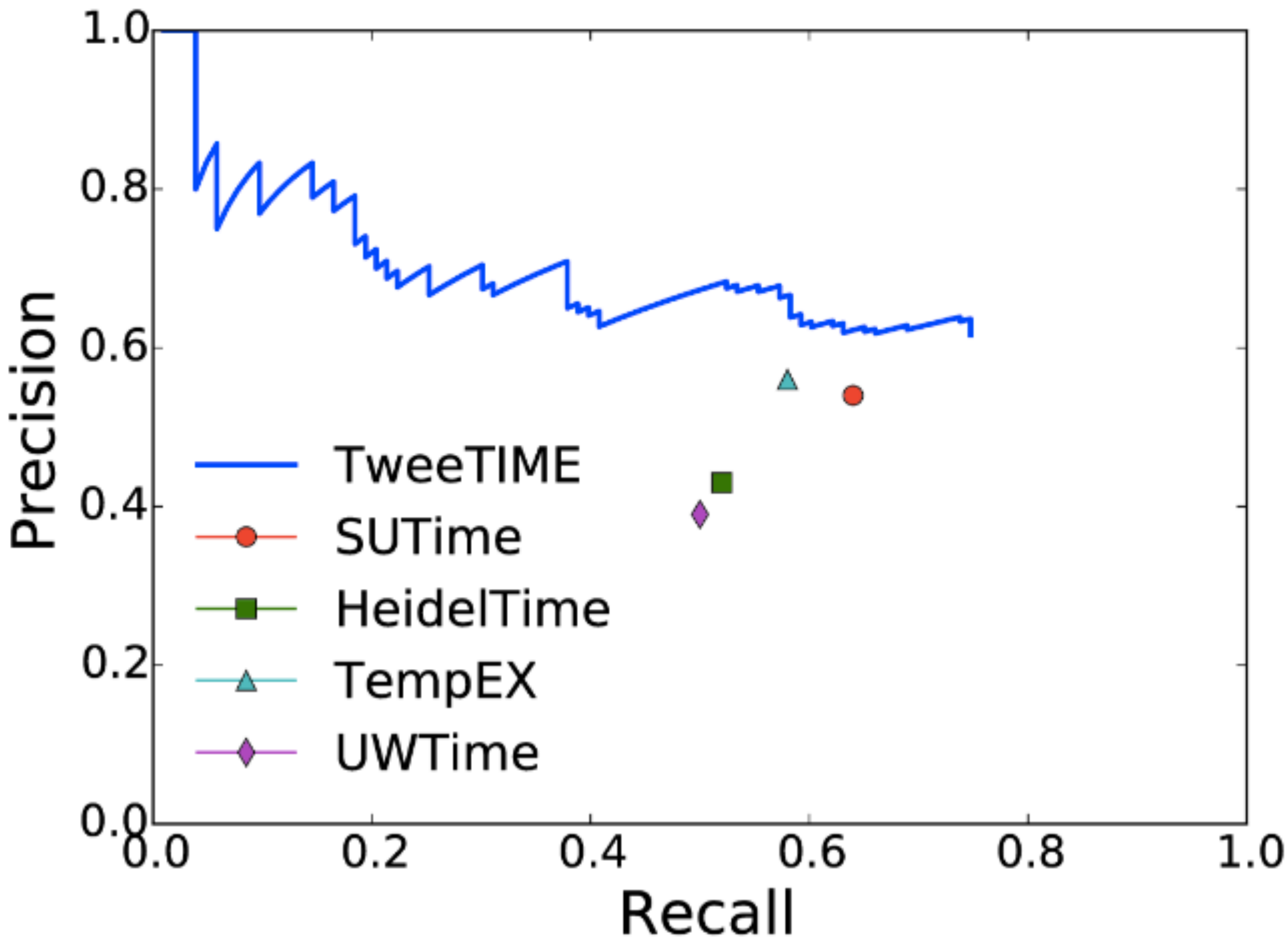
$$P(d_i|t_j, d_c) \propto e^{\theta \cdot \phi_{ij}(t_j, d_c, tags_j)}$$

D_{-10}	...	D_{-1}	D_0	D_{+1}	...	D_{+10}	$None$
10/22		10/31	11/1	11/2		11/11	NA

Temporal Normalizer

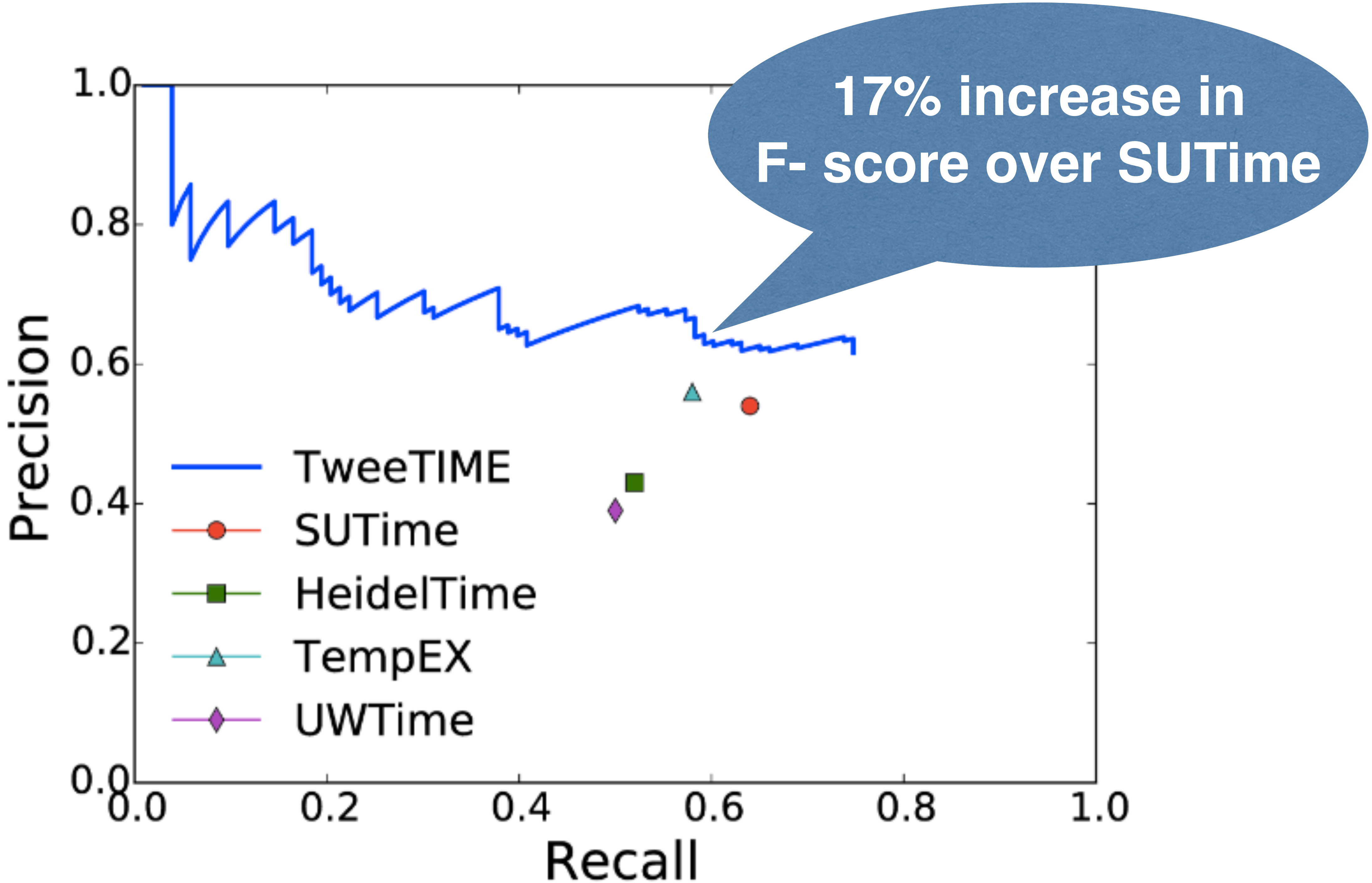


Evaluation



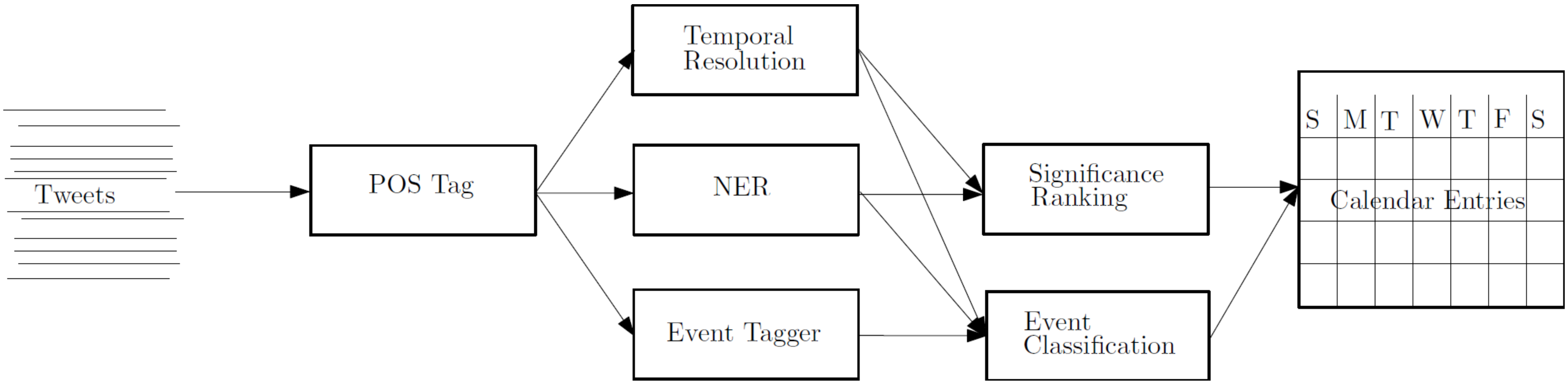
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
TweeTIME
SUTime
HeidelTime
TempEX
UWTime

Evaluation



17% increase in
F- score over SUTime
1.0
0.8
0.6
0.4
0.2
0.0
Precision
0.0
0.2
0.4
0.6
0.8
1.0
Recall
TweeTIME
SUTime
HeidelTime
TempEX
UWTime

Extracting a Calendar from Twitter



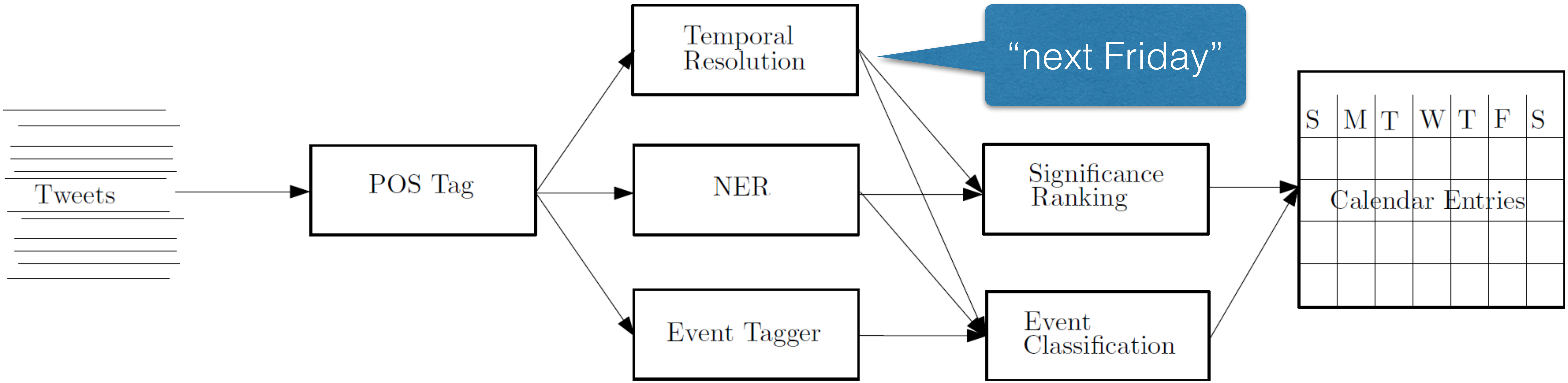
- Stream of 2 million Tweets per day
- In-domain NLP tools:

  https://github.com/aritter/twitter_nlp

- **Demo:** http://statuscalendar.com

[Ritter, et. al. KDD 2012]

Extracting a Calendar from Twitter



- Stream of 2 million Tweets per day
- In-domain NLP tools:

  https://github.com/aritter/twitter_nlp

- **Demo:** http://statuscalendar.com

[Ritter, et. al. KDD 2012]

Extracting a Calendar from Twitter

May 2016

Mon	Tue	Wed	Thu	Fri
23	24	25	26	27
republican: vote, closed primary, register california: vote, register, deadline	adwords: happen, confirms, shows google: confirms, happen, shows analytics: join, announce, held the last blade: strikes, coming, launches	itunes allowances: end, sharing, ending support dc universe: comes out, launches, begins #waywardpines: season, moving, premiere richard	john wayne day: declare, scrapped, rejected club ivy: | info sawyer fredericks: winner, live, returns agnez mo: empire tnw conference europe 2016:	alice: looking, comes out, fall big board game day: free, sign up, sponsoring dj hurrikane: hosted, coming, @imepromotions dj specks: hosted, coming,

count: 292
score: 1.7/100

"California #Republicans, vote for Hillary in the #CAPrimary! Re-register by May 23rd. #GOP #NeverTrump

[Ritter, et. al. KDD 2012]

Extracting a Calendar from Twitter

May 2016

Mon	Tue	Wed	Thu	Fri
23	24	25	26	27
republican: vote, closed primary, register california: vote, register, deadline count: 292 score: 1.7/100 "California #Republicans, vote for Hillary in the #CAPrimary! Re-register by May 23rd. #GOP #NeverTrump	adwords: happen, confirms, shov goog conf happ analytics: join, announce, held the last blade: strikes, coming, launches	itunes allowances: end, sharing, ng ort iverse: s out, hes, begins #waywardpines: season, moving, premiere richard	john wayne day: declare, scrapped, rejected club ivy: | info sawyer fredericks: winner, live, returns agnez mo: empire tnw conference europe 2016:	alice: looking, comes out, fall big board game day: free, sign up, sponsoring dj hurrikane: hosted, coming, @imepromotions dj specks: hosted, coming,

[Ritter, et. al. KDD 2012]

Extracting a Calendar from Twitter

June 2016

Wed	Thu	Fri	Sat	Sun
15	16	17	18	19
billy welu: application deadline premier league: released, season, breaking jonathan ramsden: coo, step down, #trading summer lunch 15 june:	#ki2016: #marketing, connecting, speaking breakfast flights emirates: de: going, going to, @atlasgenius dover: going, going to, @atlasgenius	wwdc: held, according, hit count: 81 score: 12.34/100 "Apple's WWDC will be held June 13 to June 17 #strategysmedia #socialmedia https://t.co/JAB6LXxcs7" meeting, @volbeat, @realkingdiamond	indian air force: pilot, join, empowerment arup raha: pilot, chief, flying dessel: meeting, @kseofficial, @nightwishband ca: going to, announced,	worlds vital #exhibition #jozi 1 march: mmvas: take place, confirmed, live karura forest: join, check out, date @mastercarduk: announce, delighted, take place

[Ritter, et. al. KDD 2012]

Extracting a Calendar from Twitter

June 2016

Wed	Thu	Fri	Sat	
15	16	17	18	
billy welu: application deadline premier league: released, season, breaking jonathan ramsden: coo, step down, #trading summer lunch 15 june:	#ki2016: #marketing, connecting, speaking breakfast flights emirates: de: going, going to, @atlasgenius dover: going, going to, @atlasgenius	wwdc: held, according, hit score: 12.34/100 "Apple's WWDC will be held June 13 to June 17 #strategysmedia #socialmedia https://t.co/JAB6LXxcs7" meeting, @volbeat, @realkingdiamond	indian air force: pilot, join, empowerment arup raha: pilot, chief, flying dessel: meeting, @kseofficial, @nightwishband ca: going to, announced,	mmvas: take place, confirmed, live karura forest: join, check out, date @mastercarduk: announce, delighted, take place

Uncertain Outcomes

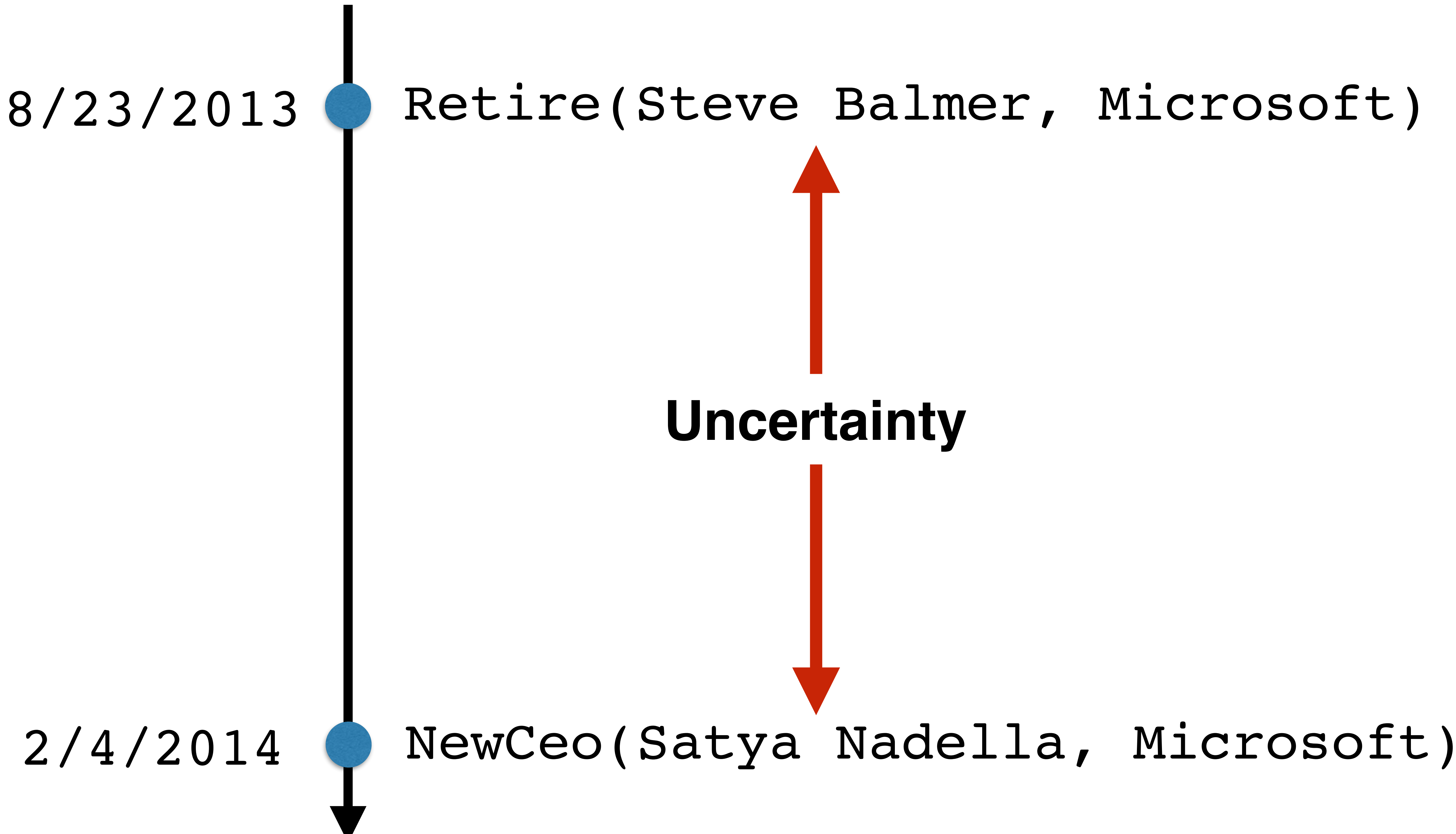
8/23/2013 Retire(Steve Balmer, Microsoft)

Uncertain Outcomes

8/23/2013 Retire(Steve Balmer, Microsoft)

2/4/2014 NewCeo(Satya Nadella, Microsoft)

Uncertain Outcomes



Uncertain Outcomes

8/23/2013 `Retire(Steve Balmer, Microsoft)`

Follow

My bet is that Satya Nadella is named new $MSFT CEO on Ballmer's retirement.

9:32 AM - 23 Aug 2013

2/4/2014 `NewCeo(Satya Nadella, Microsoft)`

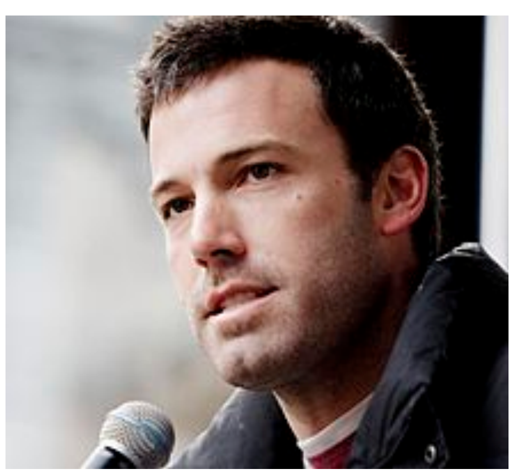
Uncertain Outcomes

8/23/2013 Retire(Steve Balmer, Microsoft)

Follow

My bet is that Satya Nadella is named new $MSFT CEO on Ballmer's retirement.

9:32 AM - 23 Aug 2013



Follow

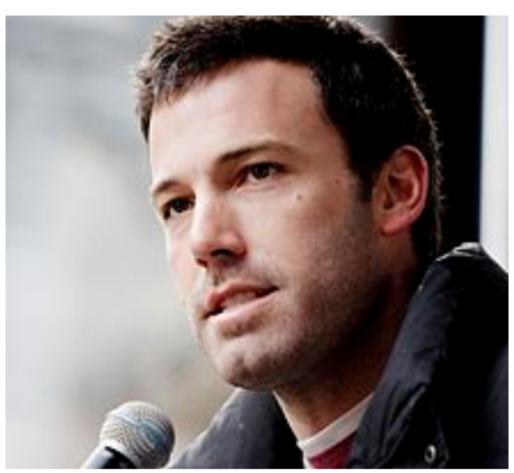
I hope Ben Affleck is the new Microsoft CEO

7:59 PM - 23 Aug 2013

2/4/2014 NewCeo(Satya Nadella, Microsoft)

Uncertain Outcomes

8/23/2013 Retire(Steve Balmer, Microsoft)



Follow

My bet is that Satya Nadella is named new $MSFT CEO on Ballmer's retirement.

9:32 AM - 23 Aug 2013

Follow

I hope Ben Affleck is the new Microsoft CEO

7:59 PM - 23 Aug 2013

2/4/2014 NewCeo(Satya Nadella, Microsoft)

Text-Driven Forecasting

(Smith et. al. 2010)

$$T_{t-1} \xrightarrow{\text{Predict}} M_t$$

Text written at time t-1

Measurement at time t

Text-Driven Forecasting

(Smith et. al. 2010)

$$T_{t-1} \xrightarrow{\text{Predict}} M_t$$

Text written at time t-1 | Measurement at time t

Examples:

Movie Reviews $\longrightarrow$ Gross Revenues

(Joshi et. al. 2010)

Text-Driven Forecasting

(Smith et. al. 2010)

$$T_{t-1} \xrightarrow{\text{Predict}} M_t$$

Text written at time t-1 — Measurement at time t

Examples:

Movie Reviews → Gross Revenues

(Joshi et. al. 2010)

Microblogs → Influenza Rates

(Paul et. al. 2014)

Text-Driven Forecasting

(Smith et. al. 2010)

$$T_{t-1} \xrightarrow{\text{Predict}} M_t$$

Text written at time t-1 — Measurement at time t

Examples:

Movie Reviews → Gross Revenues

(Joshi et. al. 2010)

Microblogs → Influenza Rates

(Paul et. al. 2014)

Writing Style → Book Sales

(Ashok et. al. 2013)

Text-Driven Forecasting

(Smith et. al. 2010)

$$T_{t-1} \xrightarrow{\text{Predict}} M_t$$

Text written at time t-1

Measurement at time t

Fit Historical Data

$$\min_{\theta} \left[\sum_{t=1}^{N} \left(M_t - \theta \cdot f(T_{t-1}, D_{t-1}) \right)^2 \right]$$

Text-Driven Forecasting

(Smith et. al. 2010)

$$T_{t-1} \xrightarrow{\text{Predict}} M_t$$

Text written at time t-1

Measurement at time t

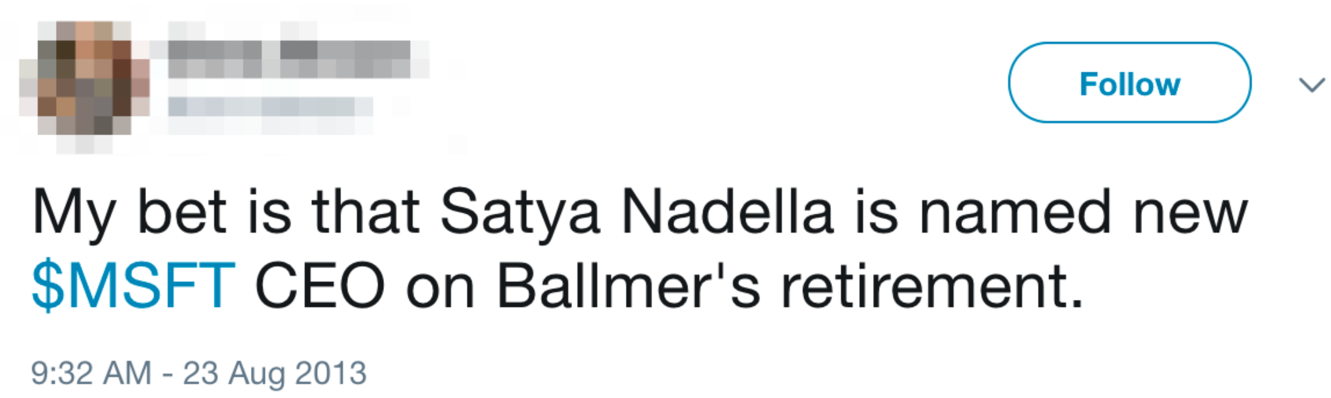
Fit Historical Data

$$\min_{\theta} \left[\sum_{t=1}^{N} \left(M_t - \theta \cdot f(T_{t-1}, D_{t-1}) \right)^2 \right]$$

Limitation: This only works for narrow domains.

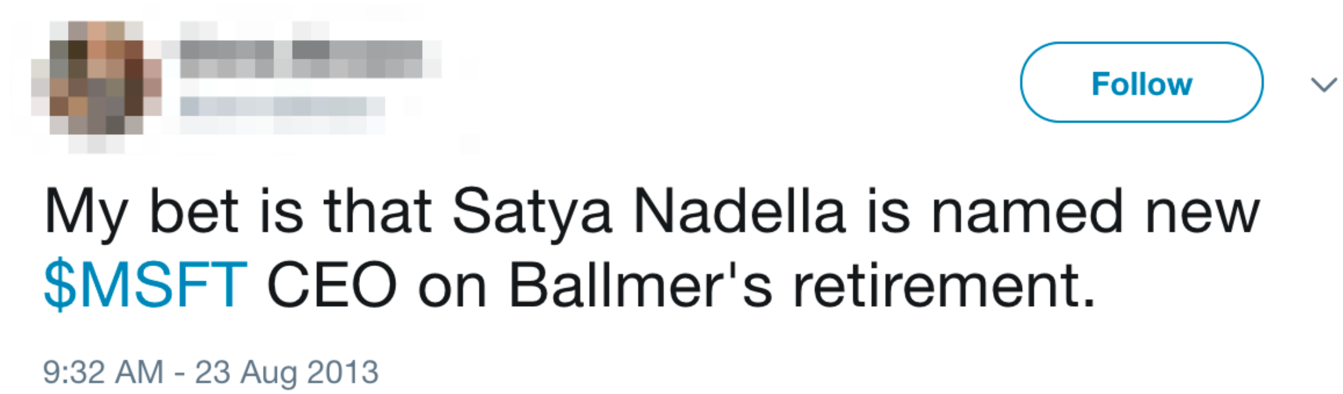
Advantages of User Predictions

CEO:



Advantages of User Predictions

CEO:



Oscars:

Advantages of User Predictions

CEO:

Follow

My bet is that Satya Nadella is named new $MSFT CEO on Ballmer's retirement.

9:32 AM - 23 Aug 2013

Oscars:

Follow

I feel that Leonardo DiCaprio is gonna win this Oscar.

4:35 PM - 27 Feb 2016

Primary Elections:

Follow

Tomorrow I believe Kasich will win Ohio, Trump will win FL. While I think Cruz has a shot at all 3 IL, NC, & MO as does Trump. #SuperTuesday

7:53 PM - 14 Mar 2016

Advantages of User Predictions

CEO:

Follow

My bet is that Satya Nadella is named new $MSFT CEO on Ballmer's retirement.

9:32 AM - 23 Aug 2013

Open Domain!

No Historical Data Required

Follow

rio is gonna win

Primary Elections:

Follow

Tomorrow I believe Kasich will win Ohio, Trump will win FL. While I think Cruz has a shot at all 3 IL, NC, & MO as does Trump. #SuperTuesday

7:53 PM - 14 Mar 2016

Data Collection (Example: Oscars)

WIKIPEDIA
The Free Encyclopedia

scrape →

Leonardo DiCaprio
Bryan Cranston
Matt Daemon
Michael Fassbender
Eddie Redmayne

(Best actor, 2/28/2016)

Data Collection (Example: Oscars)

WIKIPEDIA
The Free Encyclopedia

scrape

Leonardo DiCaprio
Bryan Cranston
Matt Daemon
Michael Fassbender
Eddie Redmayne

(Best actor, 2/28/2016)

query

Oscars **Leonardo DiCaprio** win since:2016-2-27 until:2016-2-28

Oscars **Bryan Cranston** win since:2016-2-27 until:2016-2-28

⋮

Data Collection (Example: Oscars)

WIKIPEDIA
The Free Encyclopedia

scrape

Leonardo DiCaprio
Bryan Cranston
Matt Daemon
Michael Fassbender
Eddie Redmayne

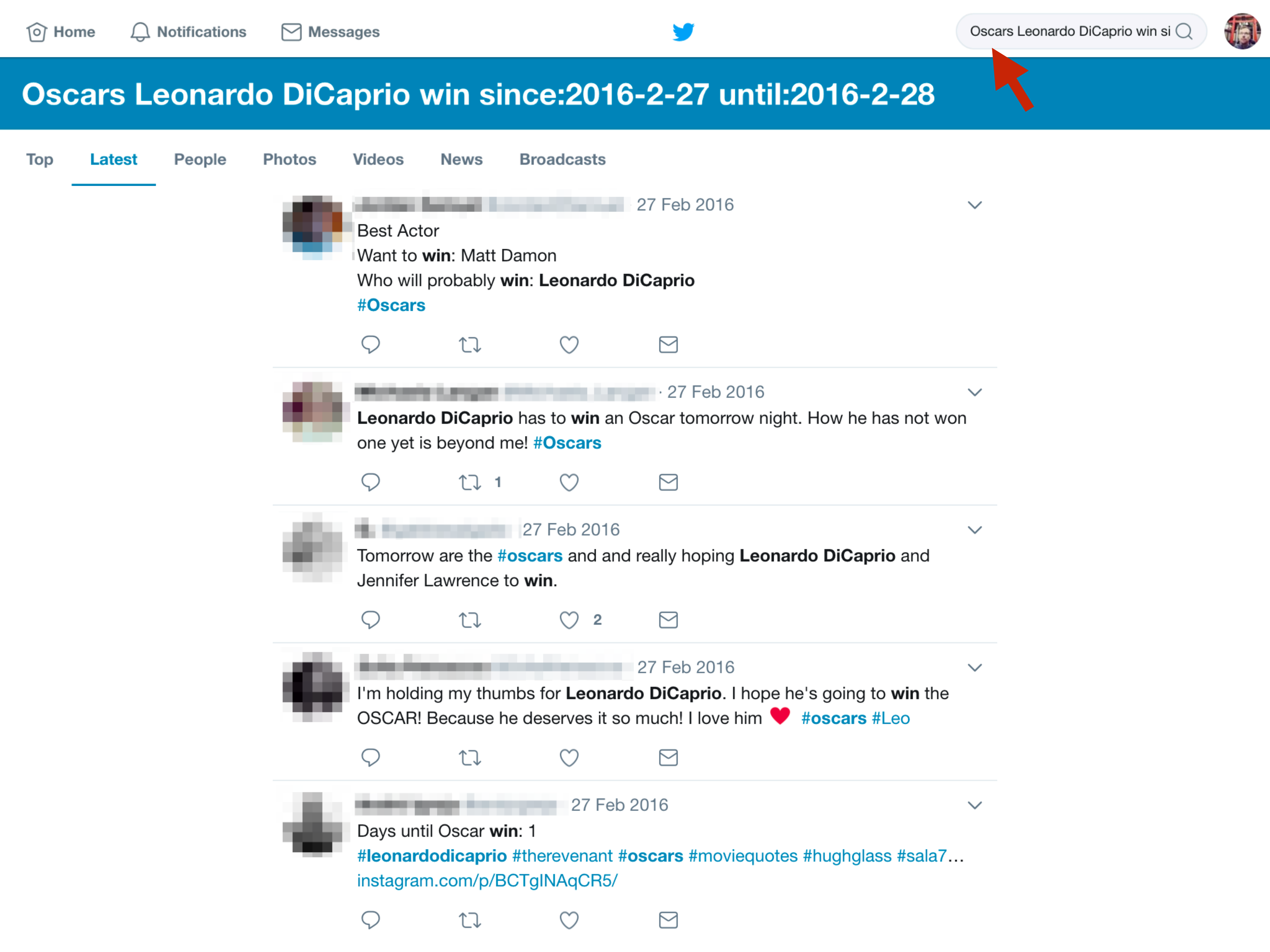
(Best actor, 2/28/2016)

query

Oscars **Leonardo DiCaprio** win since:2016-2-27 until:2016-2-28

Oscars **Bryan Cranston** win since:2016-2-27 until:2016-2-28

⋮


Home
Notifications
Messages
Oscars Leonardo DiCaprio win si
Oscars Leonardo DiCaprio win since:2016-2-27 until:2016-2-28
Top
Latest
People
Photos
Videos
News
Broadcasts
27 Feb 2016
Best Actor
Want to **win**: Matt Damon
Who will probably **win**: **Leonardo DiCaprio**
#**Oscars**
27 Feb 2016
Leonardo DiCaprio has to **win** an Oscar tomorrow night. How he has not won one yet is beyond me! #**Oscars**
1
27 Feb 2016
Tomorrow are the #**oscars** and and really hoping **Leonardo DiCaprio** and Jennifer Lawrence to **win**.
2
27 Feb 2016
I'm holding my thumbs for **Leonardo DiCaprio**. I hope he's going to **win** the OSCAR! Because he deserves it so much! I love him ❤ #**oscars** #Leo
27 Feb 2016
Days until Oscar **win**: 1
#**leonardodicaprio** #therevenant #**oscars** #moviequotes #hughglass #sala7…
instagram.com/p/BCTgINAqCR5/

Data Collection (Summary)

Contest	#Events	#Tweets
2016 US Presidential Primaries	483	38,220
Tennis Grand Slams	52	30,530
2016 US Presidential Elections	76	13,645
Football World Cup (2010-2016)	12	13,618
Balon d'Or Award (2010-2016)	18	6,777
Cricket World Cup (2010-2016)	18	4,120
Oscars (2009-2016)	176	2,370
Eurovision (2010-2016)	162	1,682
2014 Indian General Elections	68	1,656
Rugby World Cup (2010-2016)	6	651
Total	**1071**	**113269**

Data Collection (Summary)

Contest	#Events	#Tweets
2016 US Presidential Primaries	483	38,220
Tennis Grand Slams	52	30,530
2016 US Presidential Elections	76	13,645
Football World Cup (2010-2016)	12	13,618
Balon d'Or Award (2010-2016)	18	6,777
Cricket World Cup (2010-2016)	18	4,120
Oscars (2009-2016)	176	2,370
Eurovision (2010-2016)	162	1,682
2014 Indian General Elections	68	1,656
Rugby World Cup (2010-2016)	6	651
Total	**1071**	**113269**

- **Not all are Positive Veridicality Predictions**
- **We Need a Classifier to Measure Veridicality**

Mechanical Turk Annotation

Tweet - "Leonardo DiCaprio will win at the Oscars! Best Performance ever!"

a. *Based on the tweet above, does **the author** think that*

Leonardo DiCaprio is going to win at the Oscars?

- (•) Definitely Yes
- ( ) Probably Yes
- ( ) The author is uncertain about the outcome
- ( ) Probably No
- ( ) Definitely No

b. *What is the **author's desire** towards*

Leonardo DiCaprio winning at the Oscars

- ( ) Strongly wants the event to happen
- ( ) Probably wants the event to happen
- (•) No desire about the event
- ( ) Probably does not want the event to happen
- ( ) Strongly against the event

Mechanical Turk Annotation

Tweet - "Leonardo DiCaprio will win at the Oscars! Best Performance ever!"

a. *Based on the tweet above, does **the author** think that*

Leonardo DiCaprio is going to win at the Oscars?

- (●) Definitely Yes
- ( ) Probably Yes
- ( ) The author is uncertain about the outcome
- ( ) Probably No
- ( ) Definitely No

b. *What is the **author's desire** towards*

Leonardo DiCaprio winning at the Oscars

- ( ) Strongly wants the event to happen
- ( ) Probably wants the event to happen
- (●) No desire about the event
- ( ) Probably does not want the event to happen
- ( ) Strongly against the event

Mechanical Turk Annotation

Tweet - "Leonardo DiCaprio will win at the Oscars! Best Performance ever!"

a. *Based on the tweet above, does* ***the author*** *think that*

Leonardo DiCaprio is going to win at the Oscars?

- Definitely Yes
- Probably Yes
- The author is uncertain about the outcome
- Probably No
- Definitely No

b. *What is the* ***author's desire*** *towards*

Leonardo DiCaprio winning at the Oscars

- Strongly wants the event to happen
- Probably wants the event to happen
- No desire about the event
- Probably does not want the event to happen
- Strongly against the event

Mechanical Turk Annotation

Tweet - "Leonardo DiCaprio will win at the Oscars! Best Performance ever!"

a. *Based on the tweet above, does **the author** think that*

Leonardo DiCaprio is going to win at the Oscars?

- Definitely Yes
- Probably Yes
- The author is uncertain about the outcome
- Probably No
- Definitely No

b. *What is the **author's desire** towards*

Leonardo DiCaprio winning at the Oscars

- Strongly wants the event to happen
- Probably wants the event to happen
- No desire about the event
- Probably does not want the event to happen
- Strongly against the event

Mechanical Turk Annotation

Tweet - "Leonardo DiCaprio will win at the Oscars! Best Performance ever!"

a. *Based on the tweet above, does **the author** think that*

Leonardo DiCaprio is going to win at the Oscars?

- (●) Definitely Yes
- ( ) Probably Yes
- ( ) The author is uncertain about the outcome
- ( ) Probably No
- ( ) Definitely No

b. *What is the **author's desire** towards*

Leonardo DiCaprio winning at the Oscars

- ( ) Strongly wants the event to happen
- ( ) Probably wants the event to happen
- (●) No desire about the event
- ( ) Probably does not want the event to happen
- ( ) Strongly against the event

- 3,543 Tweets
- Each tweet annotated by 7 AMT Workers
- MACE **(Hovy et. al. 2017)** was used to resolve differences

Measuring the Veridicality of Users' Predictions

$$\log P(y = v | c, O, \text{tweet}) \propto \theta_v \cdot f(c, O, \text{tweet})$$

Veridicality

(positive, negative, neutral)

Measuring the Veridicality of Users' Predictions

$$\log P(y = v | c, O, \text{tweet}) \propto \theta_v \cdot f(c, O, \text{tweet})$$

Veridicality
(positive, negative, neutral)

Candidate
(e.g. Kasich)

Measuring the Veridicality of Users' Predictions

List of Opponents
(e.g. Trump, Rubio, …)

$$\log P(y = v | c, O, \mathrm{tweet}) \propto \theta_v \cdot f(c, O, \mathrm{tweet})$$

Veridicality
(positive, negative, neutral)

Candidate
(e.g. Kasich)

Features

$$f(c, O, \text{tweet})$$

Leonardo DiCaprio will *win* Oscars for Revenant. Bad luck *Cranston*!

Features

$$f(c, O, \mathrm{tweet})$$

TARGET(t) ENTITY(e) ENTITY(e) OPPONENT(o)

Leonardo DiCaprio will *win* Oscars for Revenant. Bad luck *Cranston*!

Features

$$f(c, O, \text{tweet})$$

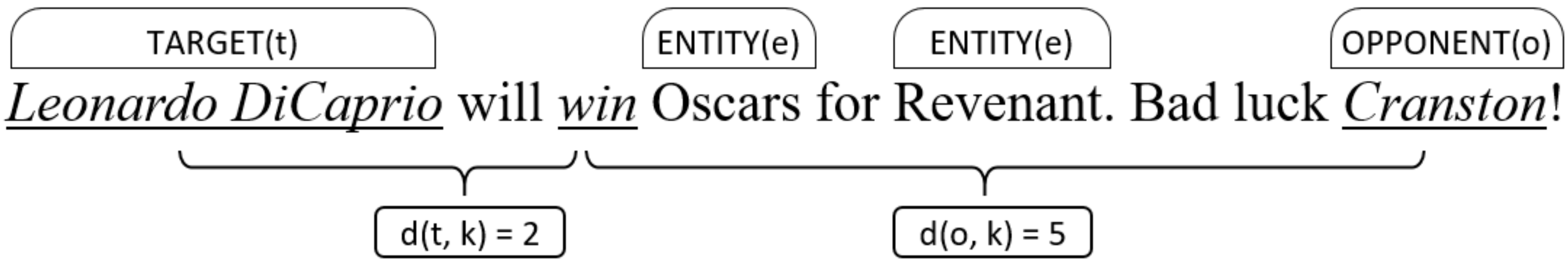
TARGET(t) ENTITY(e) ENTITY(e) OPPONENT(o)

Leonardo DiCaprio will *win* Oscars for Revenant. Bad luck *Cranston*!

Context

Features

$$f(c, O, \text{tweet})$$

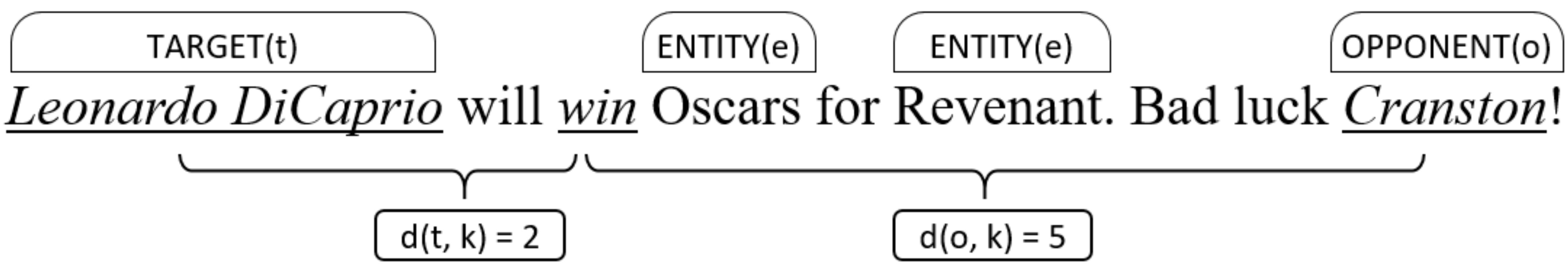


Context

Distance to Keyword

Features

$$f(c, O, \mathrm{tweet})$$



Context

Distance to Keyword

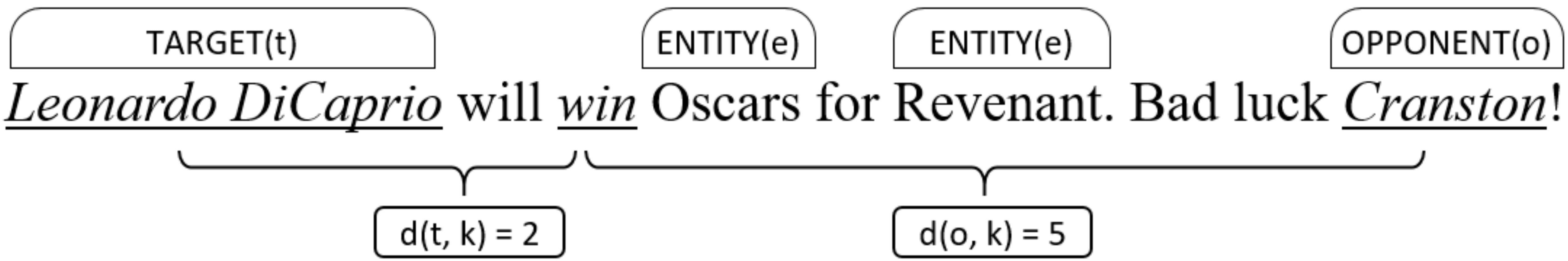
Puctuation (Exclamation marks + Question Mark)

Dependency Path Features

Negation Features

Features

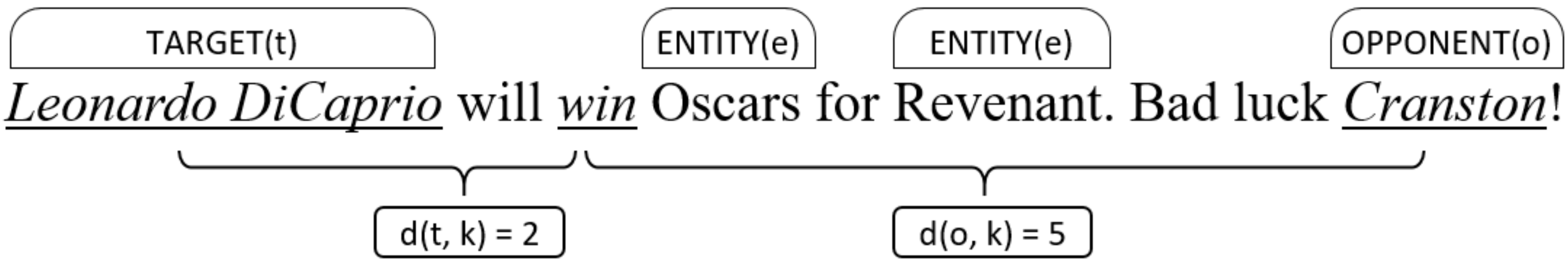
$$f(c, O, \mathrm{tweet})$$



Positive Veridicality Feature Type	Feature	Weight
Keyword context	TARGET *will* KEYWORD	0.41
Keyword dep. path	TARGET $\rightarrow$ *to* $\rightarrow$ KEYWORD	0.38
Keyword dep. path	TARGET $\leftarrow$ *is* $\rightarrow$ *going* $\rightarrow$ *to* $\rightarrow$ KEYWORD	0.29
Target context	TARGET *is favored to win*	0.19
Keyword context	TARGET *are going to* KEYWORD	0.15
Target context	TARGET *predicted to win*	0.13
Pair context	TARGET1 *could win* TARGET2	0.13
Distance to keyword	TARGET *closer to* KEYWORD	0.11

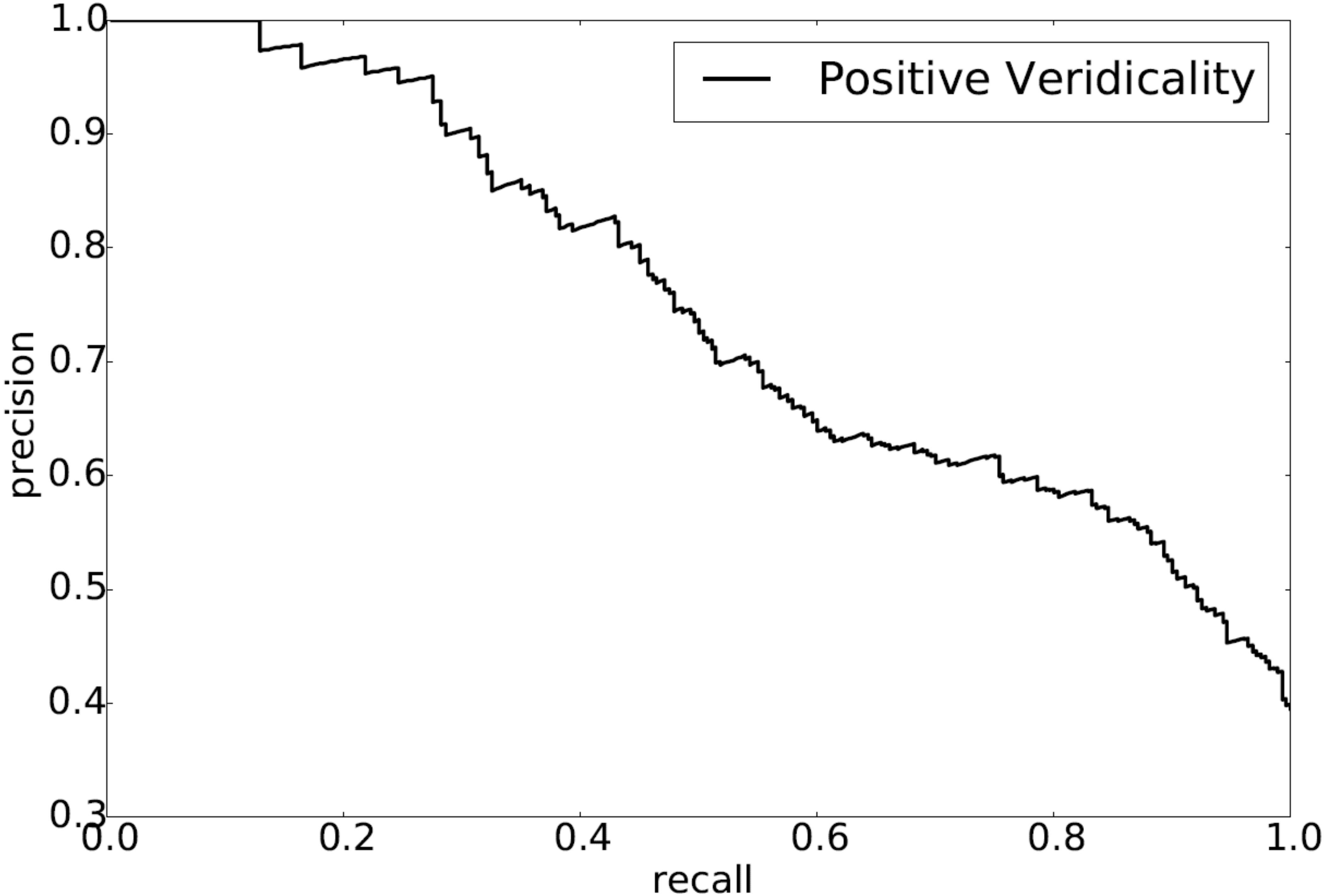
Features

$$f(c, O, \text{tweet})$$

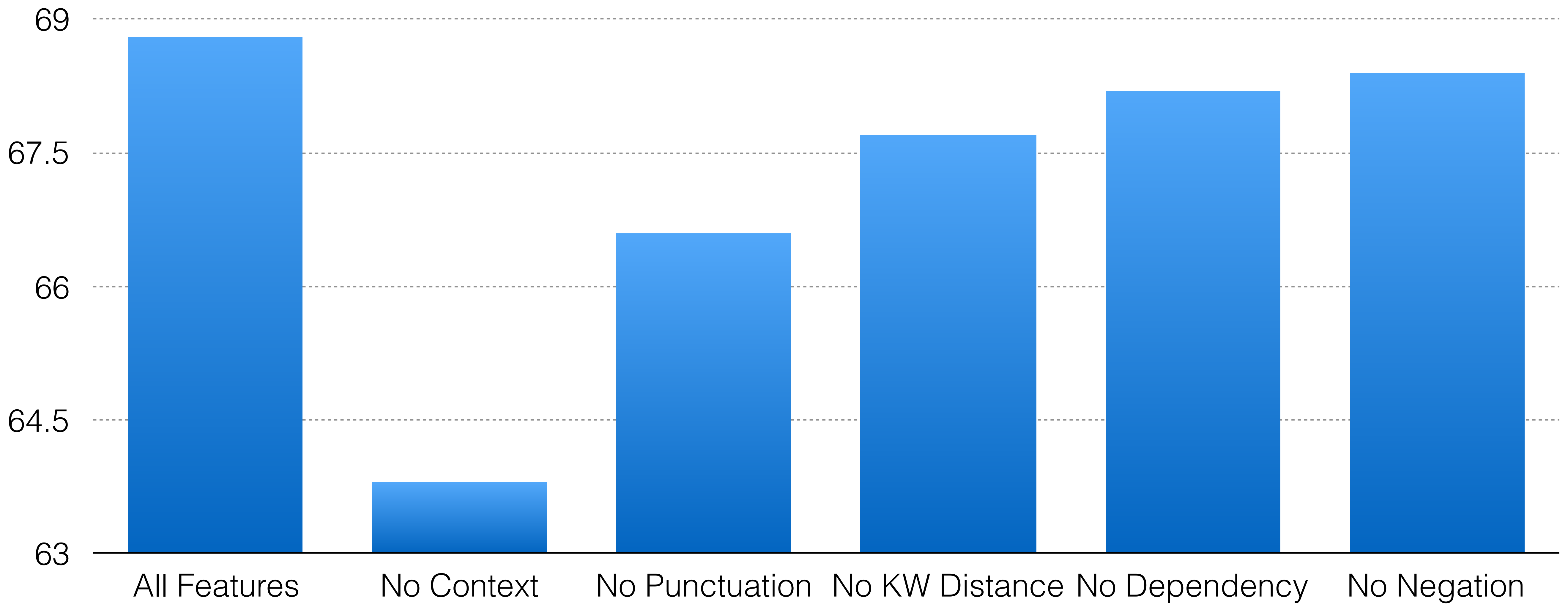


Negative Veridicality Feature Type	Feature	Weight
Negated keyword	keyword is negated	0.47
Keyword context	TARGET *won't* KEYWORD	0.41
Opponent context	OPPONENT *will win*	0.37
Keyword dep. path	TARGET $\leftarrow$ *will* $\rightarrow$ *not* $\rightarrow$ KEYWORD	0.31
Distance to keyword	OPPONENT *closer to* KEYWORD	0.28
Target context	TARGET *may not win*	0.27
Keyword dep. path	OPPONENT $\leftarrow$ *will* $\rightarrow$ KEYWORD	0.23
Target context	TARGET *can't win*	0.18

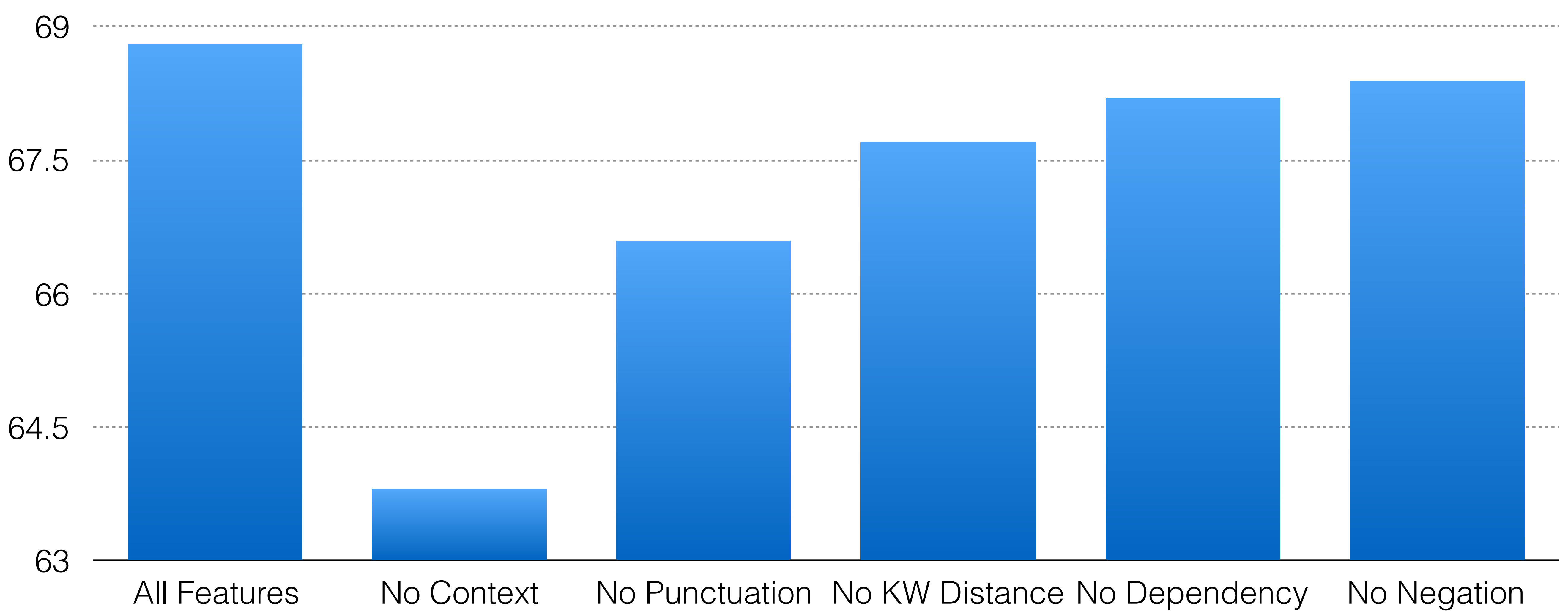
Precision / Recall



Feature Ablation (Max F1)

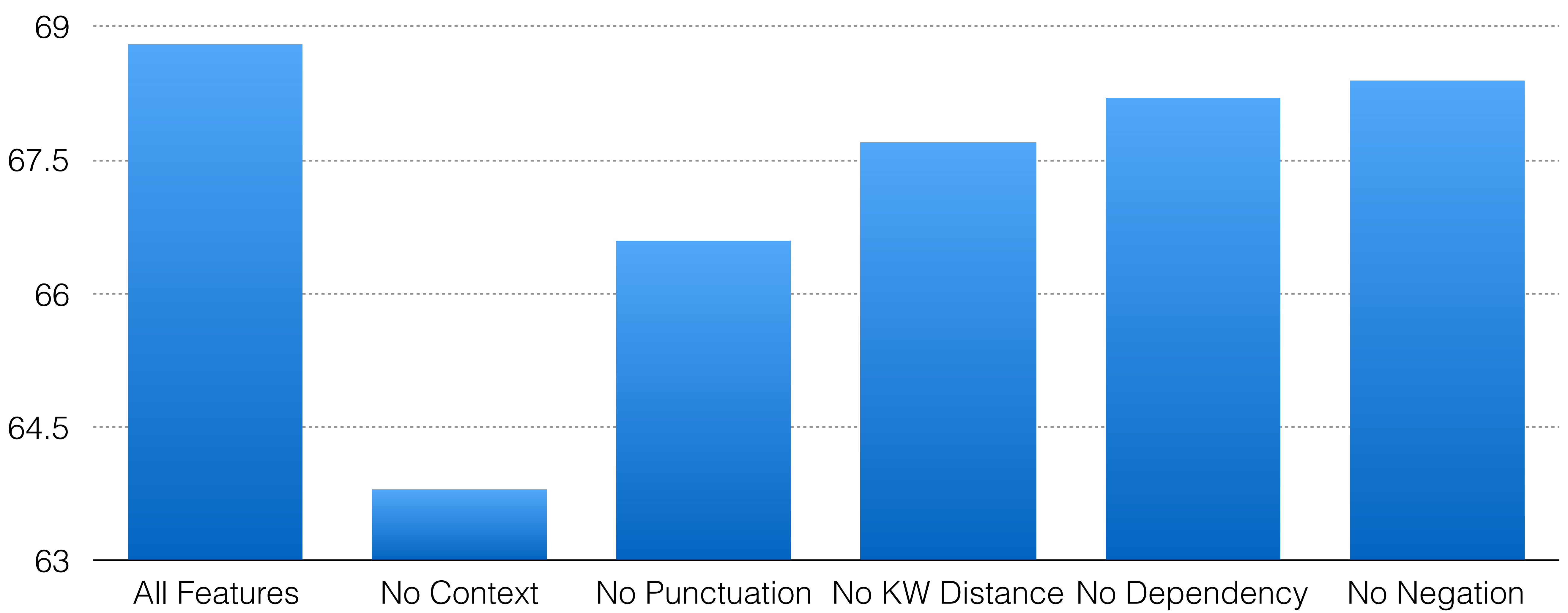


Feature Ablation (Max F1)



Context features are crucial

Feature Ablation (Max F1)



Dependency features help a bit

Error Analysis

Error Analysis

Tweet	Gold	Predicted
The heart wants **Nadal** to win tomorrow but the mind points to a Djokovic win over 4 sets. Djokovic 7-5 4-6 7-5 6-4 **Nadal** for me.	?	

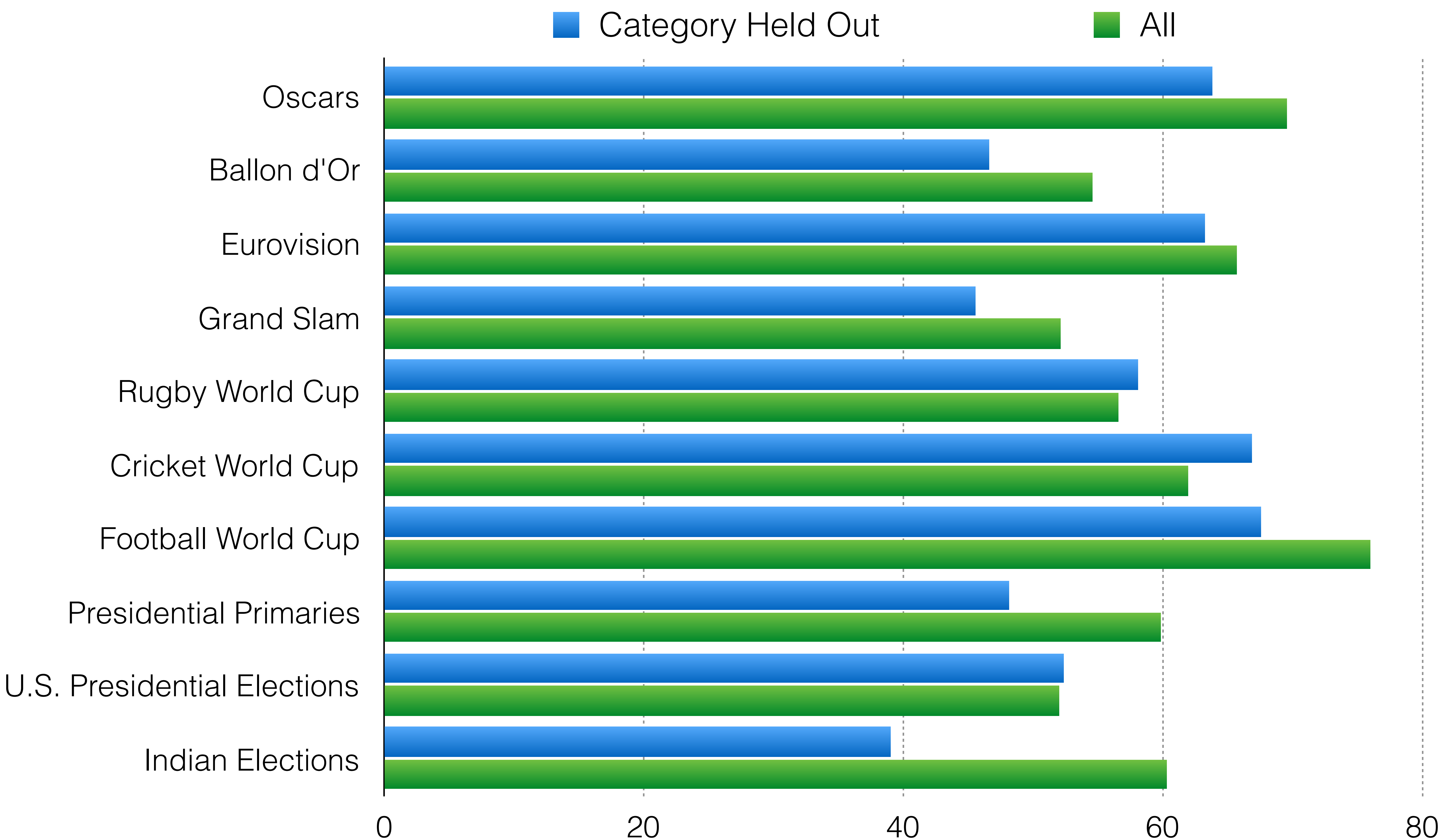
Error Analysis

Tweet	Gold	Predicted
The heart wants **Nadal** to win to-morrow but the mind points to a Djokovic win over 4 sets. Djokovic 7-5 4-6 7-5 6-4 **Nadal** for me.	negative	positive

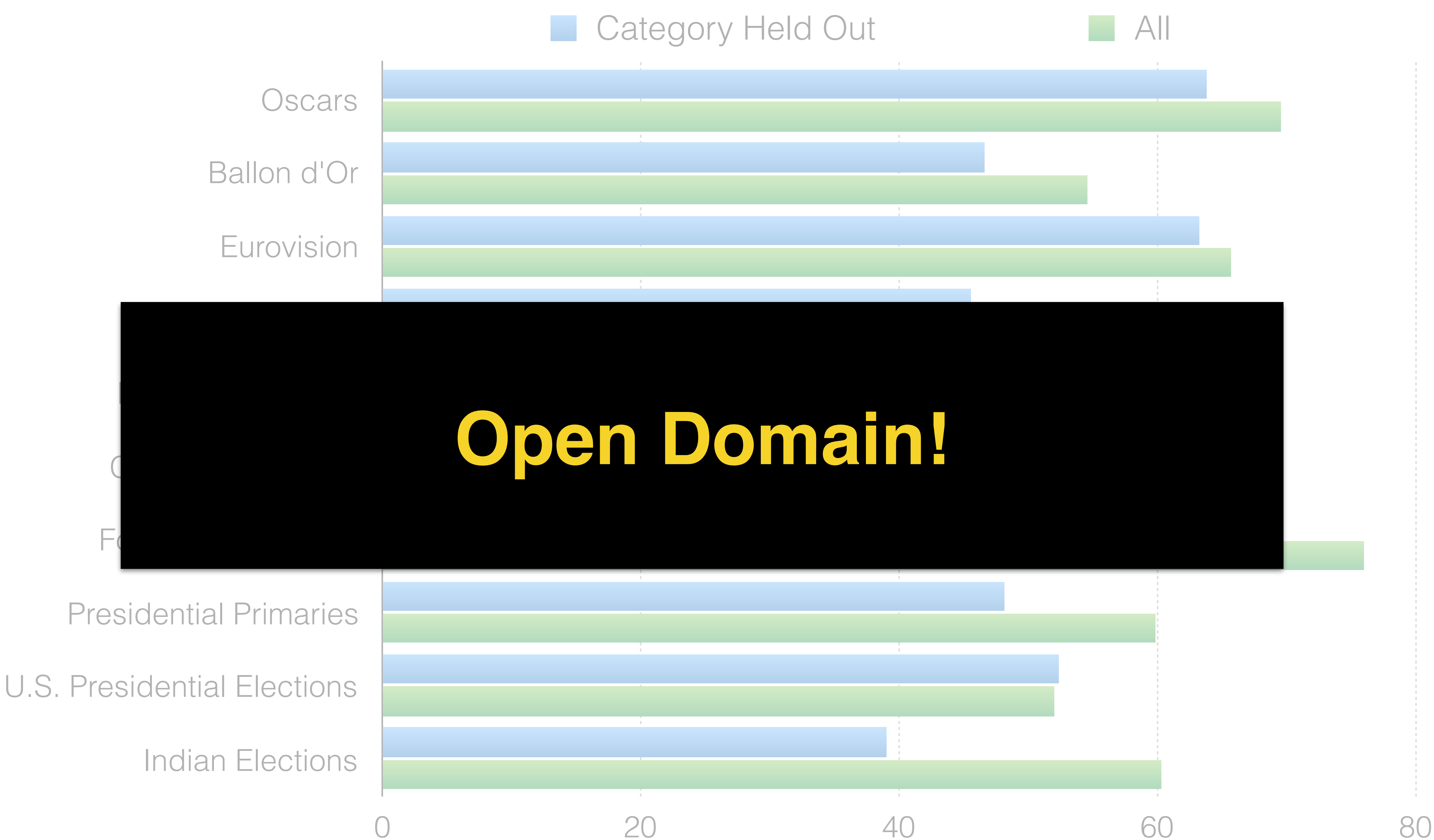
Error Analysis

Tweet	Gold	Predicted
The heart wants **Nadal** to win tomorrow but the mind points to a Djokovic win over 4 sets. Djokovic 7-5 4-6 7-5 6-4 **Nadal** for me.	negative	positive
Hopefully tomorrow Federer will win and beat that **Nadal** guy lol	neutral	negative

Cross-Domain Experiments



Cross-Domain Experiments



Forecasting

Follow

My bet is that Satya Nadella is named new $MSFT CEO on Ballmer's retirement.

9:32 AM - 23 Aug 2013

Follow

If I am MSFT, Tony Bates makes a lot of sense as the next CEO. ibtimes.com/steve-ballmer- …

1:08 PM - 27 Aug 2013

Forecasting

Follow

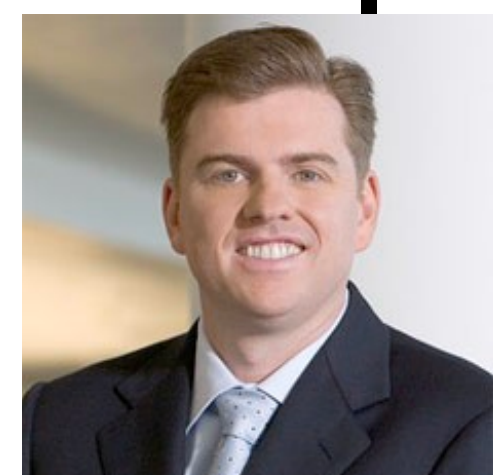
My bet is that Satya Nadella is named new $MSFT CEO on Ballmer's retirement.

9:32 AM - 23 Aug 2013

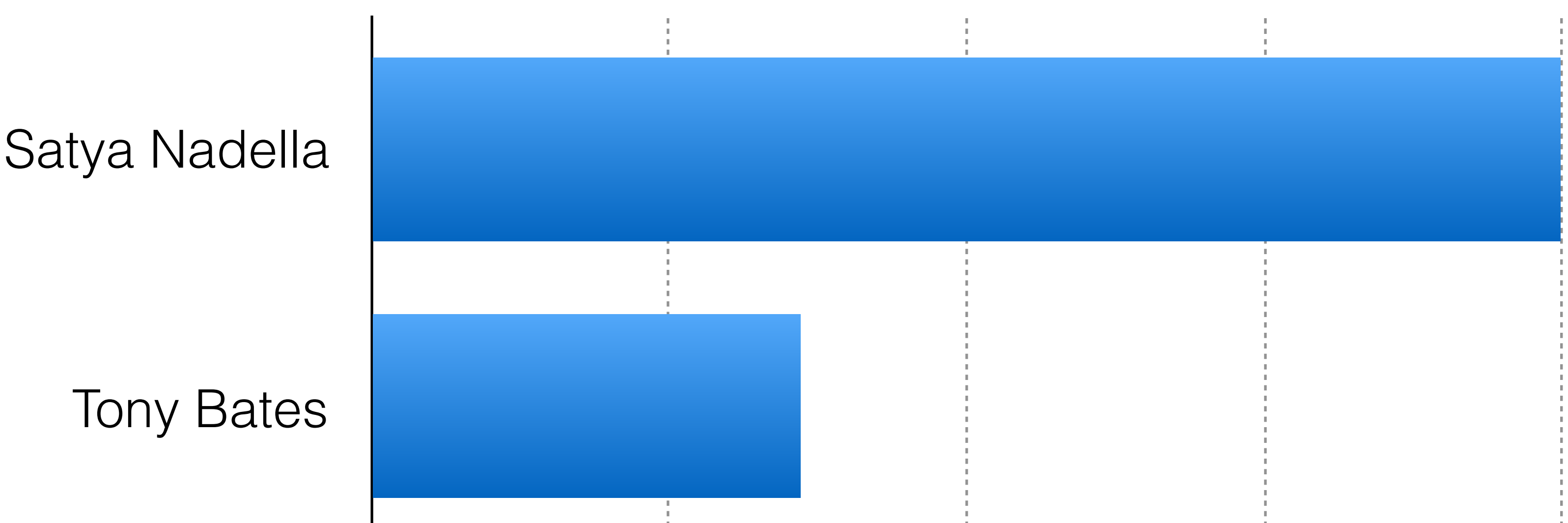
Follow

If I am MSFT, Tony Bates makes a lot of sense as the next CEO. ibtimes.com/steve-ballmer- …

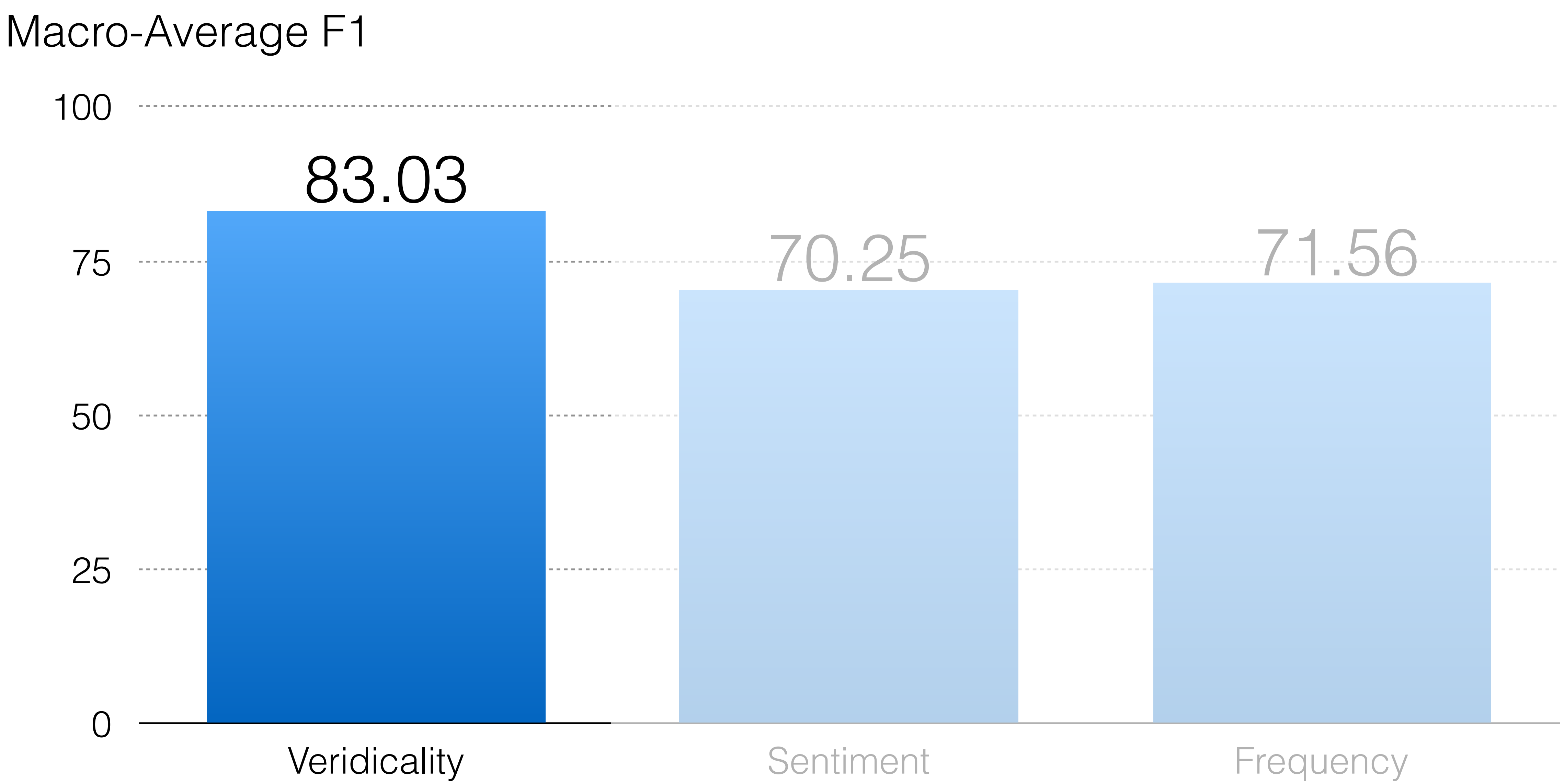
1:08 PM - 27 Aug 2013



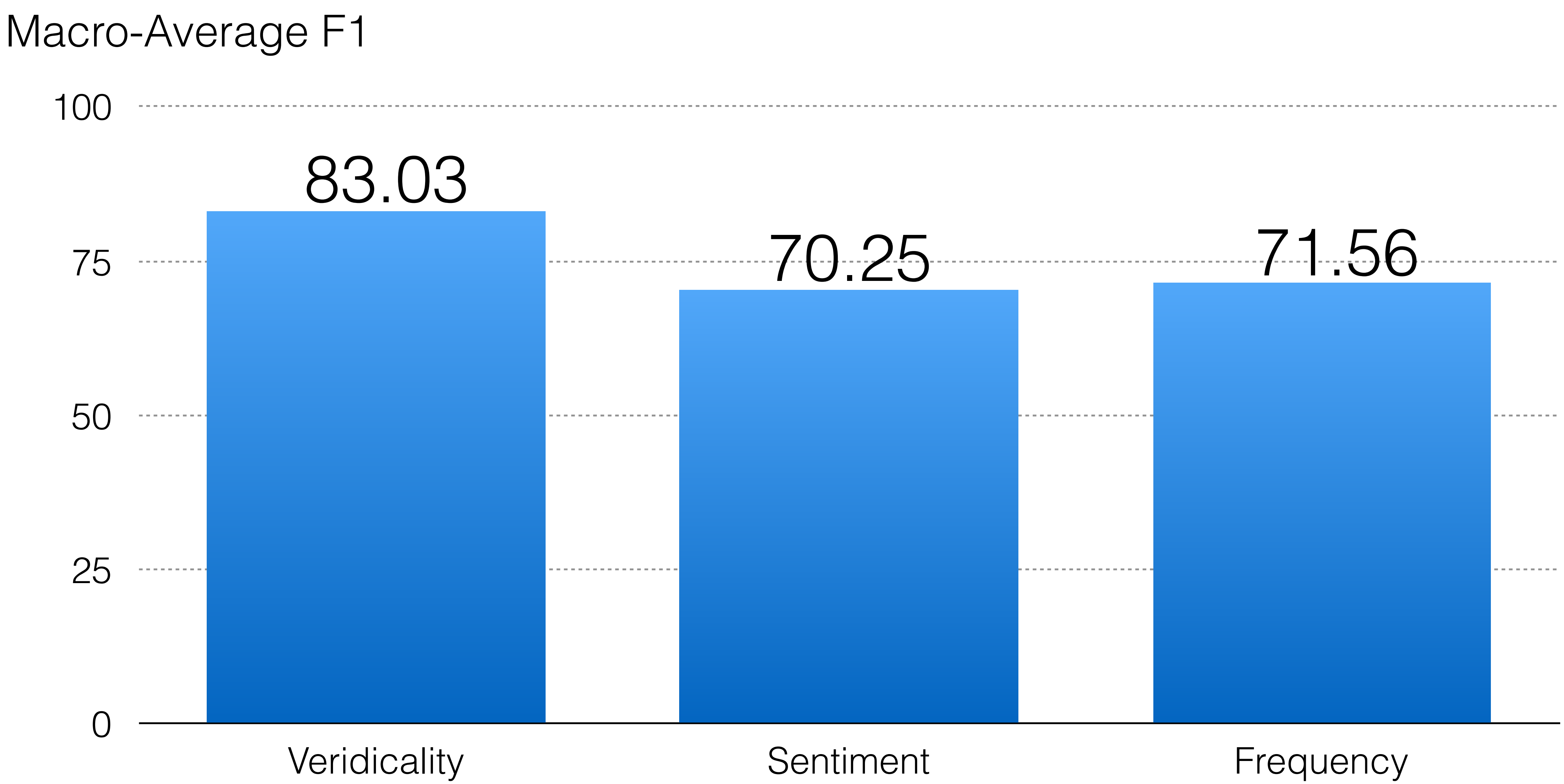
Vote



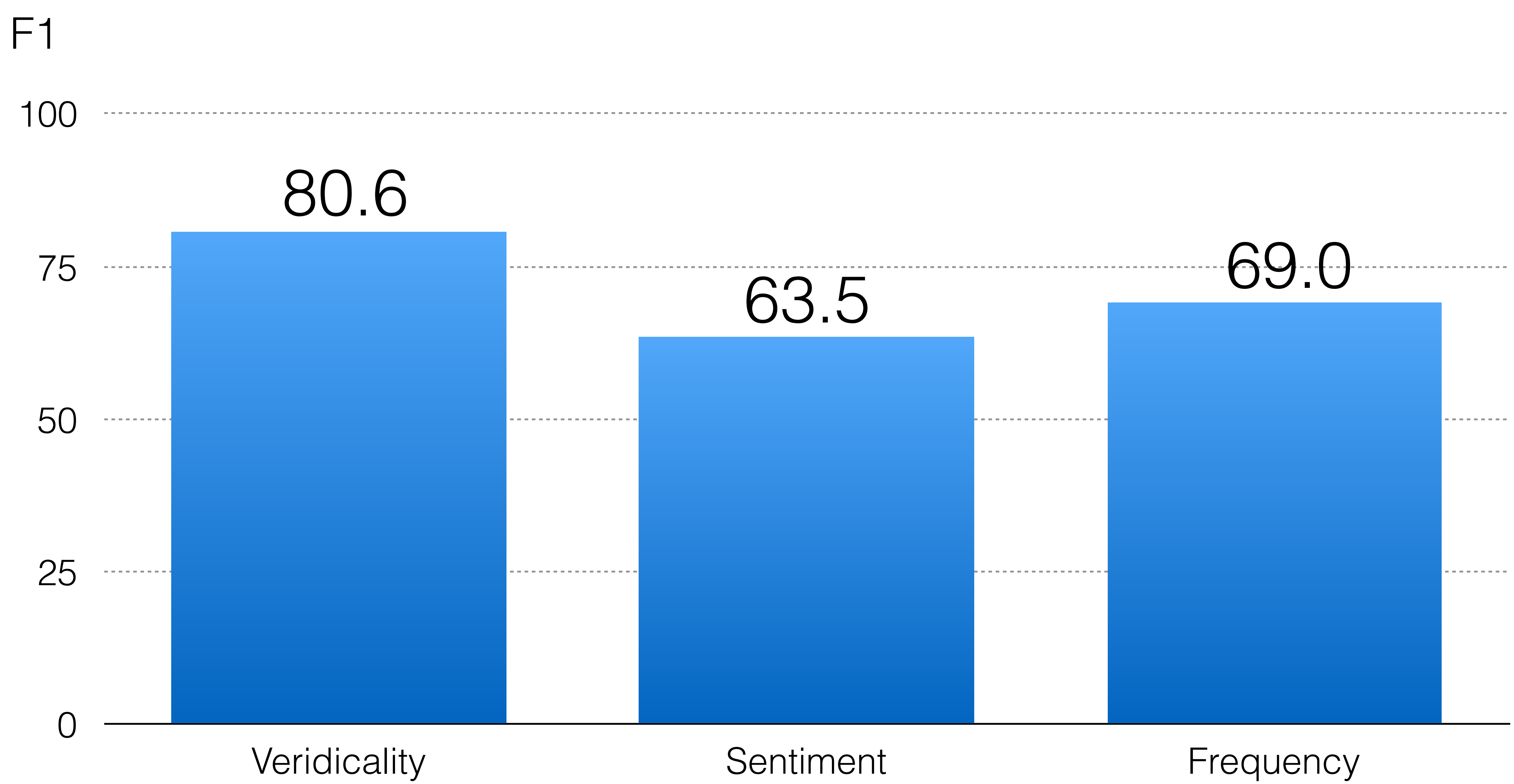
Forecasting Performance



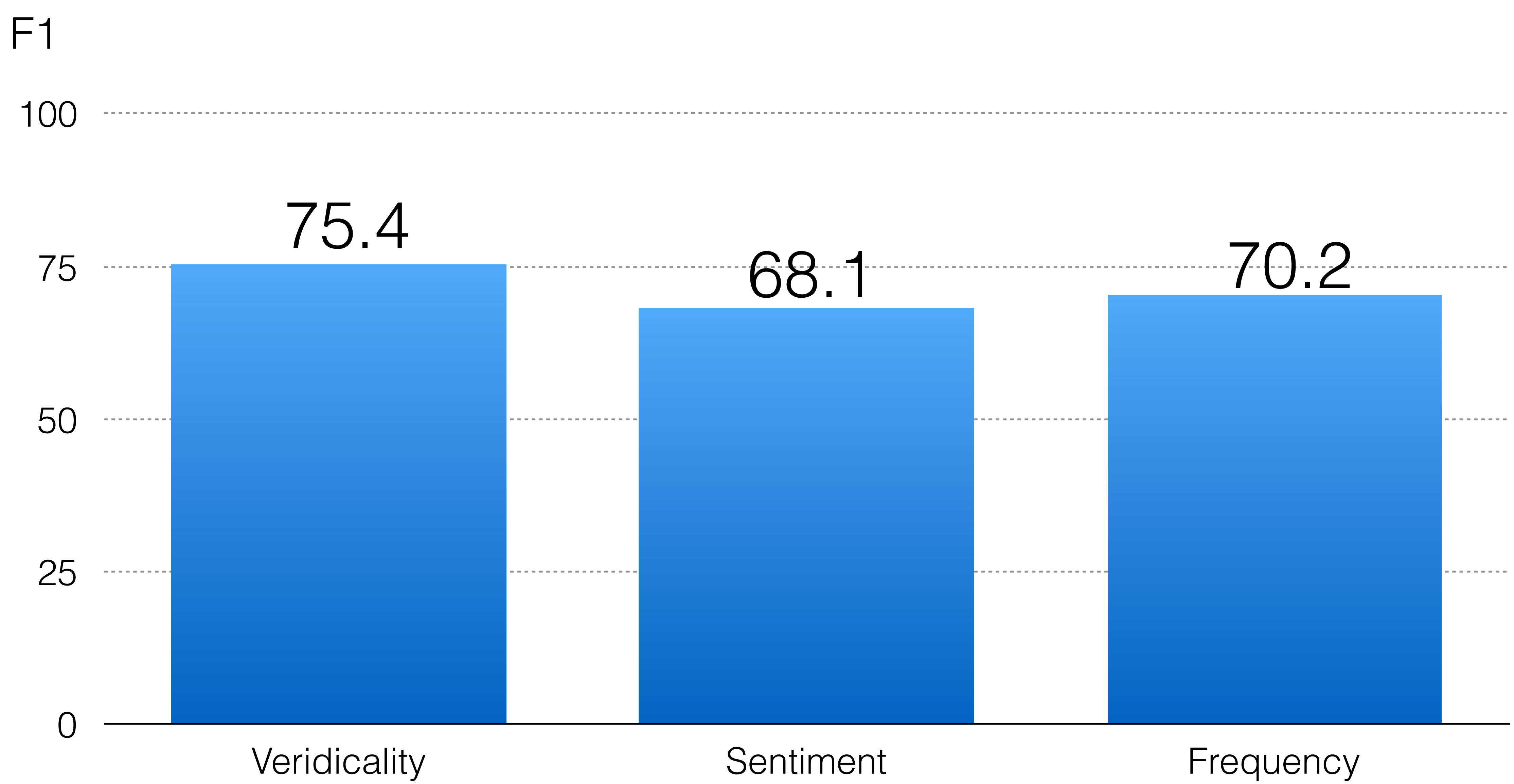
Baselines (Sentiment + Volume)



Oscars (151 Predictions)



U.S. Presidential Primaries (211 Predictions)



Surprise Outcomes

		Contender		Veridicality Score
OSCARS		**Leonardo DiCaprio**		0.97
		Natalie Portman		0.92
		Julianne Moore		0.91
		Daniel Day-Lewis		0.90
		Slumdog Millionaire		0.75
		Matthew McConaughey		0.74
	!	The Revenant		0.73
		Argo		0.71
		Brie Larson		0.70
		The Artist		0.67
PRIMARIES		**Trump**	South Carolina	0.96
		Clinton	Iowa	0.90
		Trump	Massachusetts	0.88
		Trump	Tennessee	0.88
		Sanders	Maine	0.87
		Sanders	Alaska	0.87
	!	Trump	Maine	0.87
		Sanders	Wyoming	0.86
		Trump	Louisiana	0.86
	!	Clinton	Indiana	0.85

Surprise Outcomes

		Contender		Veridicality Score
OSCARS		**Leonardo DiCaprio**		0.97
		Natalie Portman		0.92
		Julianne Moore		0.91
		Daniel Day-Lewis		0.90
		Slumdog Millionaire		0.75
		Matthew McConaughey		0.74
	!	The Revenant		0.73
		Argo		0.71
		Brie Larson		0.70
		The Artist		0.67
PRIMARIES		**Trump**	South Carolina	0.96
		Clinton	Iowa	0.90
		Trump	Massachusetts	0.88
		Trump	Tennessee	0.88
		Sanders	Maine	0.87
		Sanders	Alaska	0.87
	!	Trump	Maine	0.87
		Sanders	Wyoming	0.86
		Trump	Louisiana	0.86
	!	Clinton	Indiana	0.85


The Washington Post

Arts and Entertainment

How ‘Spotlight’s’ surprise Best Picture win made Oscar history

By Amy Argetsinger and Geoff Edgers February 29, 2016

Surprise Outcomes

		Contender		Veridicality Score
OSCARS		**Leonardo DiCaprio**		0.97
		Natalie Portman		0.92
		Julianne Moore		0.91
		Daniel Day-Lewis		0.90
		Slumdog Millionaire		0.75
		Matthew McConaughey		0.74
	!	The Revenant		0.73
		Argo		0.71
		Brie Larson		0.70
		The Artist		0.67
PRIMARIES		**Trump**	South Carolina	0.96
		Clinton	Iowa	0.90
		Trump	Massachusetts	0.88
		Trump	Tennessee	0.88
		Sanders	Maine	0.87
		Sanders	Alaska	0.87
	!	Trump	Maine	0.87
		Sanders	Wyoming	0.86
		Trump	Louisiana	0.86
	!	Clinton	Indiana	0.85

The Washington Post

Arts and Entertainment

How 'Spotlight's' surprise Best Picture win made Oscar history

By Amy Argetsinger and Geoff Edgers February 29, 2016

HUFFPOST

Ted Cruz Upsets Donald Trump In Maine Republican Caucus

Cruz won despite Gov. Paul LePage's endorsement of Trump.

By Daniel Marans

Surprise Outcomes

		Contender		Veridicality Score
OSCARS		**Leonardo DiCaprio**		0.97
		Natalie Portman		0.92
		Julianne Moore		0.91
		Daniel Day-Lewis		0.90
		Slumdog Millionaire		0.75
		Matthew McConaughey		0.74
	!	The Revenant		0.73
		Argo		0.71
		Brie Larson		0.70
		The Artist		0.67
PRIMARIES		**Trump**	South Carolina	0.96
		Clinton	Iowa	0.90
		Trump	Massachusetts	0.88
		Trump	Tennessee	0.88
		Sanders	Maine	0.87
		Sanders	Alaska	0.87
	!	Trump	Maine	0.87
		Sanders	Wyoming	0.86
		Trump	Louisiana	0.86
	!	Clinton	Indiana	0.85

The Washington Post

Arts and Entertainment

How 'Spotlight's' surprise Best Picture win made Oscar history

By Amy Argetsinger and Geoff Edgers February 29, 2016

HUFFPOST

Ted Cruz Upsets Donald Trump In Maine Republican Caucus

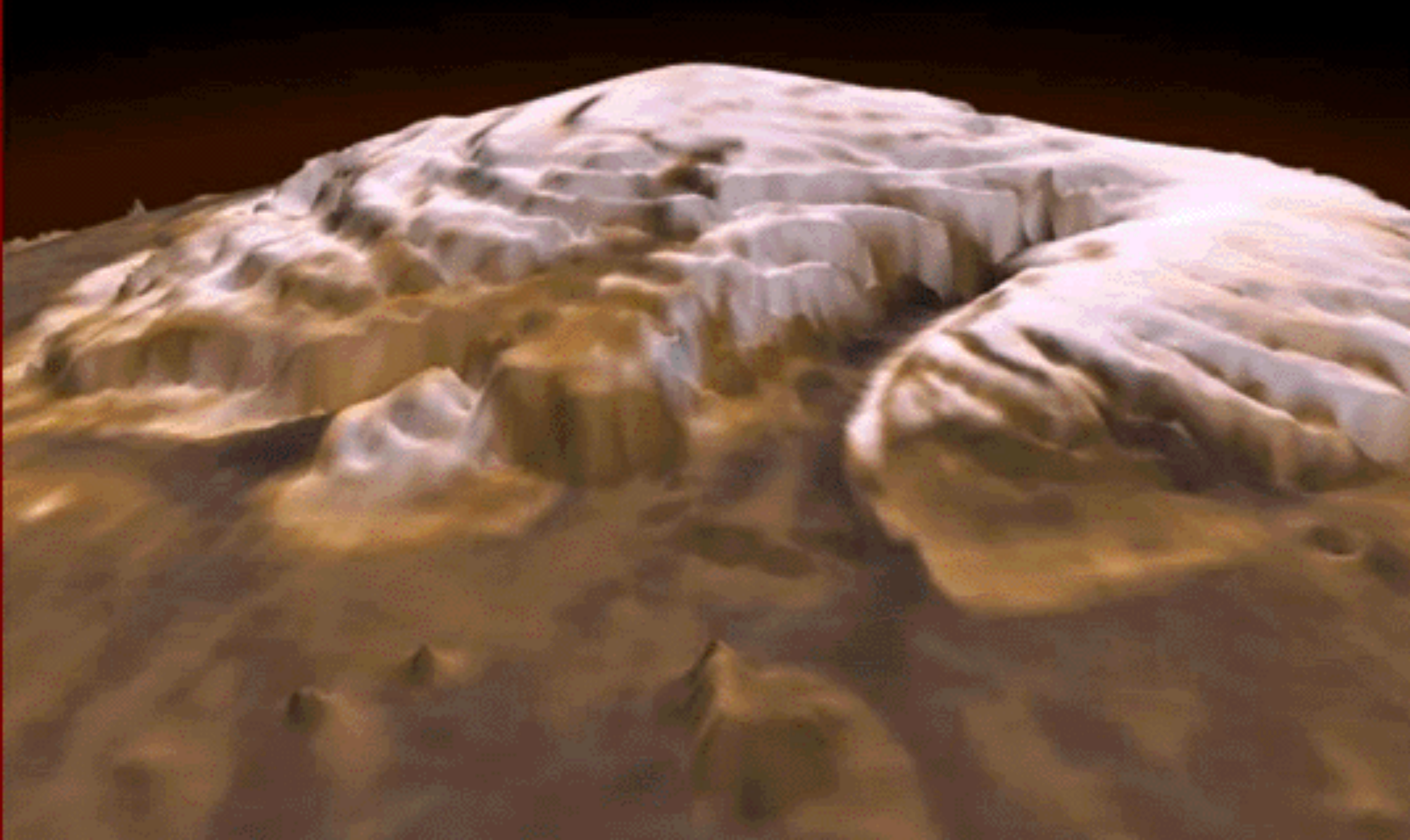
Cruz won despite Gov. Paul LePage's endorsement of Trump.

By Daniel Marans

POLITICO

Sanders surprises Clinton in Indiana

Following the Water: The Mars Exploration Program



Orlando Figueroa, Director

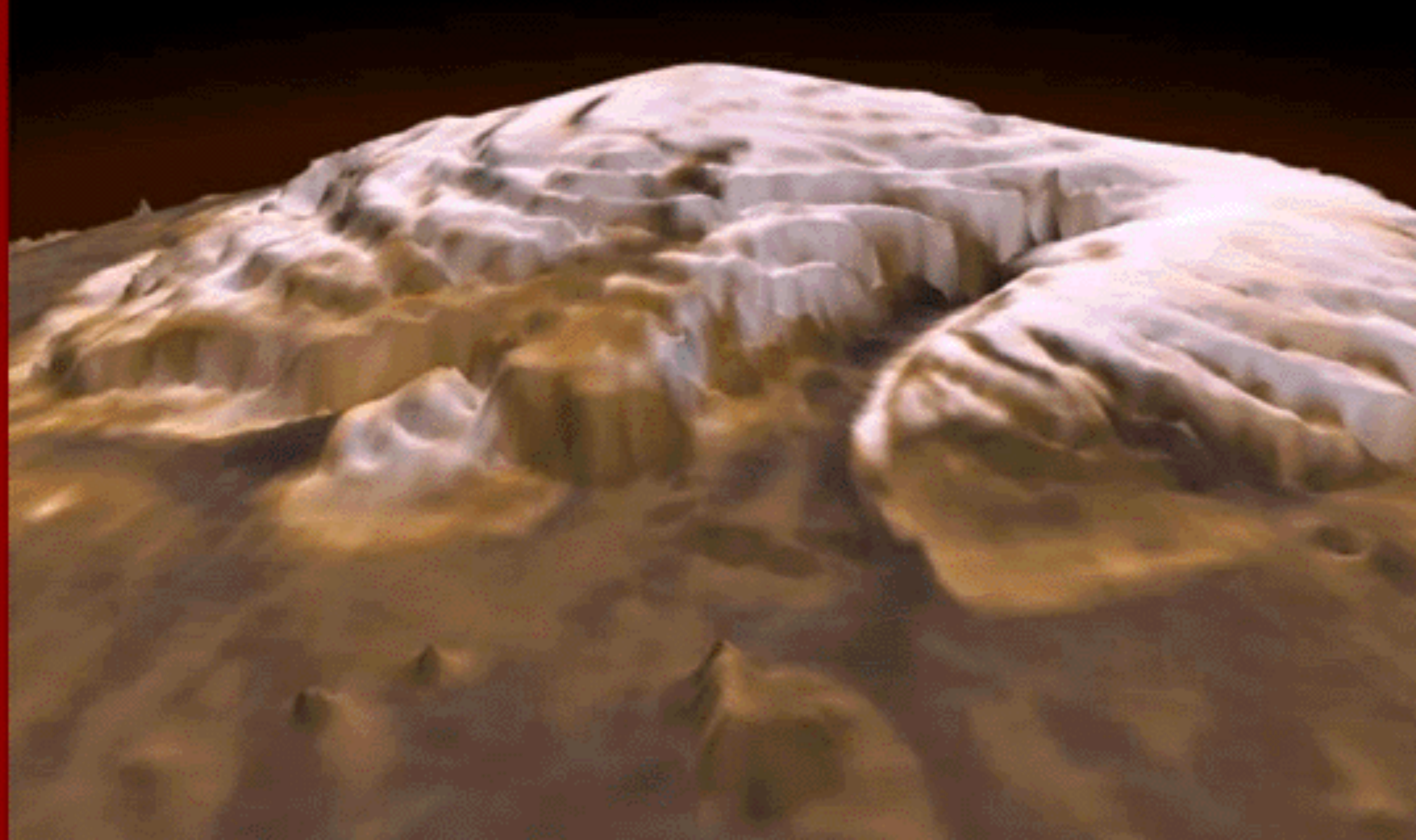
Dr. Jim Garvin, Lead Scientist

Mars Exploration Program
NASA

Following the Water:

Program

Where can we find NLU?



Orlando Figueroa,
Director

Dr. Jim Garvin,
Lead Scientist

Mars Exploration
Program
NASA

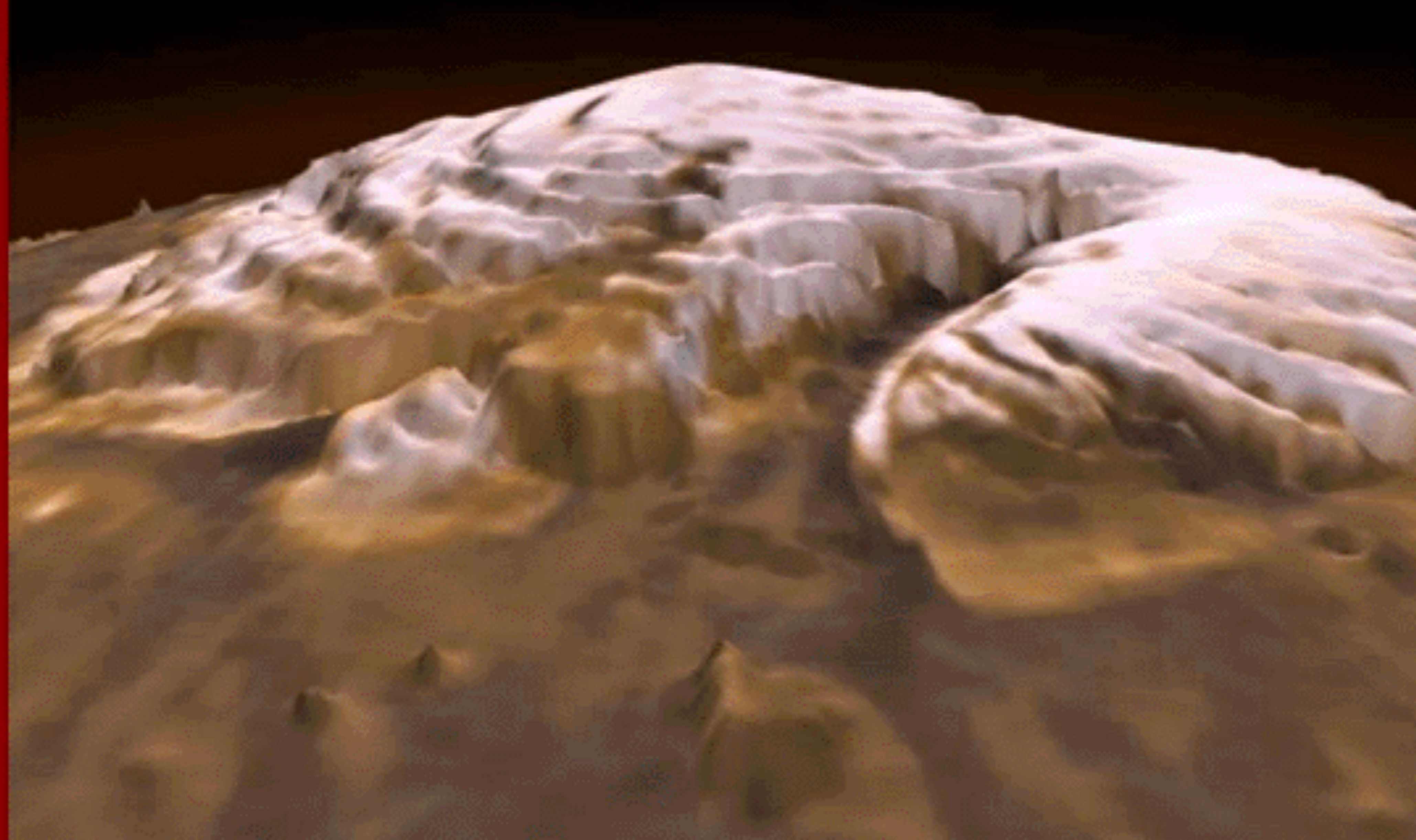
Following the Water:
Program

Where can we find NLU?
Follow the data!

Orlando Figueroa,
Director

Dr. Jim Garvin,
Lead Scientist

Mars Exploration
Program
NASA



Where can we find NLU?

Follow the data!

Opportunistically Gathered Data:

- Known Events (Forecasting, Time Normalization)
- Billions of Internet Conversations

Following the Water:

n Program

Where can we find NLU?

Follow the data!

Opportunistically Gathered Data:

- Known Events (Forecasting, Time Normalization)
- Billions of Internet Conversations

Design Models for the Data

(rather than the other way around)

Mars Exploration Program
NASA

Where can we find NLU?

Follow the data!

Opportunistically Gathered Data:

- Known Events (Forecasting, Time Normalization)
- Billions of Internet Conversations

Design Models for the Data

(rather than the other way around)

Thank You!