10 years after "I, Robot"

Is that a bot running the social media feed?
C. Edwards, A. Edwards, P. R. Spence, A. K. Shelton

Detecting Automation of Twitter Accounts
Z. Chu, S. Gianvecchio, H. Wang, S. Jajodia
What is Twitter

Marvel Entertainment
@Marvel

@starwars Can we borrow some lightsabers? We have...science...to do. #AskStarWars

Star Wars
@starwars

We'll trade you for the Mjölnir. RT @Marvel: @starwars Can we borrow some lightsabers? We have...science...to do. #AskStarWars
What is Twitter

@starwars Can we borrow some lightsabers? We have...science...to do. #AskStarWars

Star Wars
@starwars

We'll trade you for the Mjölnir. RT @Marvel: @starwars Can we borrow some lightsabers? We have...science...to do. #AskStarWars
Access
Access

Welcome to Twitter! Follow people to get great content. Std msg rates apply. Already using Twitter? Reply w/ YES to sign in. Reply w/ your full name to signup.

Congrats Larry Bird, you're now on Twitter! Your username is @LarryBi13921013. If you don't like this username, reply w/ SET USERNAME to change it.

Larry Bird
Access
Access
Access
Twitter Evolution

i can't decide if i should pursue my hopes and dreams or start watching Downton Abbey from the start of season one.

Continues heavy air strikes and counter strikes from all around Sanaa. My god. Hell is on the door. #yemen

Tunisian authorities say they've killed militant linked to #bordomuseum attack, as march against terrorism begins bbc.in/1EkjpC2
Who are the users?

Registered users: ~1 billion

Active users: ~300 millions

Idle users: ~700 millions
Who are the users?

Registered users:
~1 billion

Active users:
~300 millions

Idle users:
~700 millions
Who are the users?

HUMAN

CYBORG

BOT
@HerrMiM IT'S SHOWTIME!

@KidHannah Oh look! They've been to Saturn. Hey, I've been to Saturn! Whoa. Sandworms. You hate 'em right? I hate 'em myself!

@SharaazEasyE IT'S SHOWTIME!

@zebics IT'S SHOWTIME!

@MichaelBeek These aren't my rules. Come to think of it, I don't have any rules.

@a_dilEmma Let's turn on the juice and see what shakes loose.
@HerrMiM IT'S SHOWTIME!

@KidHannah Oh look! They've been to Saturn. Hey, I've been to Saturn! Whoa. Sandworms. You hate 'em right? I hate 'em myself!

@SharaazEasyE IT'S SHOWTIME!

@zebics IT'S SHOWTIME!

@MichaelBeek These aren't my rules. Come to think of it, I don't have any rules.

@a_dilEmma Let's turn on the juice and see what shakes loose.
Snow! Finally!

45 ore!!!

I estimated where Felix Baumgartner will land when he attempts the @redbullstratos record breaking freefall jump win.gs/wLMeKNW

I'm in a group video chat with 5 others - Click to watch or join: tinychat.com
Snow! Finally!

45 ore!!!

I estimated where Felix Baumgartner will land when he attempts the record breaking freefall jump win.gs/1JMekNw

I'm in a group video chat with 5 others - Click to watch or join: tinychat.com
Analysts suspect that Wal-Mart (WMT) was down -0.95% today because there loyalty usually resides else where.

Paychex (PAYX) lost -1.09% today amid reports that of the light pollution) but it isn’t impossible in dark suburbs.

Investors traded 2,218,708 shares of Viacom (VIAB) because I’m a miss my people’s going to Florida..

Investors traded 8,025,449 shares of Coca-Cola (KO) because MY COMPUTER’S STOPPED WORKING AND I CAN’T WATCH

MARKET REPORT: The S&P 500 lost -0.45% today because tomorrow it could be you.
Analysts suspect that Wal-Mart (WMT) was down -0.95% today because there loyalty usually resides else where.

Paychex (PAYX) lost -1.09% today amid reports that of the light pollution) but it isn't impossible in dark suburbs.

Investors traded 2,218,708 shares of Viacom (VIAB) because I'm a miss my people's going to Florida..

Investors traded 8,025,449 shares of Coca-Cola (KO) because MY COMPUTER'S STOPPED WORKING AND I CAN'T WATCH

MARKET REPORT: The S&P 500 lost -0.45% today because tomorrow it could be you.
<table>
<thead>
<tr>
<th>TWEET</th>
<th>FOLLOWING</th>
<th>FOLLOWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>460</td>
<td>108</td>
</tr>
</tbody>
</table>

Tweet

Tweet e risposte

- **dinusha @dinusha** · 19 gen 2011
  - HAHAH I JUST GOT A FREE iPHONE http://bit.ly/hd2MqQ

- **dinusha @dinusha** · 19 gen 2011

- **dinusha @dinusha** · 18 gen 2011
  - They are looking for Apple iPhone beta testers http://bit.ly/hf9nRW

- **dinusha @dinusha** · 13 dic 2010
  - Everyone should try acai! I lost 10lbs!! RT This http://to.ly/8WVB

- **dinusha @dinusha** · 12 dic 2010
  - Lose weight instantly using acai http://acainews8.com

- **dinusha @dinusha** · 28 lug 2007
  - Adding Twitter to my iPhone
SPAMBOT
Poll Time! In honour of #WorldBookDay we want to know what your favourite book was as a child....

Photo: tearyplant: tearyplant: littlealienproducts: Cat Pot Plant small // $45 plant blog plant blog tmblr.co/Zutles1lf1-7v9

@adventgeekgirl this is awesome, we love @smrtgrls! There are loads of science women in #IGGPPC so hopefully some will take part!

@IGGPPCHQ Love Science and Women? Check this out! amysmartgirls.com/2015/02/science... #iggppc #iggles
Poll Time! In honour of #WorldBookDay we want to know what your favourite book was as a child....

Photo: tearyplant: tearyplant: littlealienproducts: Cat Pot Plant small // $45
plant blog plant blog tmblr.co/Zutles1f1-7v9

@adventgeekgirl this is awesome, we love @smrtrls! There are loads of science women in IGGPPC so hopefully some will take part!

@IGGPPCHQ Love Science and Women? Check this out! amysmartgirls.com/2015/02/scienc... #iggppc #iggles
How to distinguish users?

- Friend-to-Follower ratio
- Tweeting timing
- Access patterns
- Content detection
- Account heuristics
Data Collection

DFS Crawling

Public Timeline

512.407 users
8.350.095 tweets
Ground Truth Creation

Manual check of logs and homepages to collect 2000 users per category
Ground Truth Creation

Intelligence
Originality
Consciousness

Unintelligent
Unoriginal
Automation
Followers, Friends

Followers

Friend
Friend-to-Follower Ratio

- **Bot**: aggressive following to gain friends
- **Human**: reciprocal relationships
Friend-to-Follower Ratio

\[
\text{Reputation} = \frac{\# \text{ followers}}{\# \text{ followers} + \# \text{ friends}}
\]

\[
\text{Taste} = \frac{1}{n} \sum_{\text{friends}} \left( \text{Reputation}(f) \right)
\]
Friend-to-Follower Ratio

Cumulative Distribution Function
Tweeting Timing

![Tweeting Timing Chart]

CDF vs Relative Entropy for Bot, Human, and Cyborg.
Tweeting Timing

Human Behavior

Irregularities

Larger Coding

High Entropy
Tweeting Timing

- High Entropy
  - Larger Coding
    - Irregularities
      - Human Behavior
Access Patterns

- Web
- Mobile
- API
- 3rd-Party
Content Detection
Content Detection

![Graph showing CDF vs External URL Ratio for Bot, Human, and Cyborg categories]
Content Detection

Malicious URLs:
decide whether a message is spam
based on blacklists
Content Detection

Spam detection:
Bayesian classifier decide whether a message is spam based on its content

\[
P(\text{spam}|M) = \frac{P(M|\text{spam})P(\text{spam})}{P(M)}
\]
Content Detection

Message is Spam

Author is a Bot
Content Detection

Message is Spam

Author is a Bot

What about non-malicious bots?
Account Heuristics
Decision Maker

- Input
- Log Processor
- Spammer Detection Component
- Account Properties Component
- Entropy Component
- Decision Maker
- Human
- Cyborg
- Bot
 polls!

Poll Time! In honour of #WorldBookDay we want to know what your favourite book was as a child.

Photo: tearyplant: tearyplant: littlealienproducts: Cat Pot Plant small // $45
plant blog plant blog tmblr.co/Zutles1f1-7v9

@adventgeekgirl this is awesome, we love @smtrgrls! There are loads of science women in #IGGPPC so hopefully some will take part!

It’s #WorldBookDay! Show us your costumes!

@IGGPPCHQ Love Science and Women? Check this out! amysmartgirls.com/2015/02/scienc... #iggppc #iggles

Photo: tmblr.co/Zutles1e-J9oc
Decision Maker

<table>
<thead>
<tr>
<th>Feature</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>82.8</td>
</tr>
<tr>
<td>URL Ratio</td>
<td>74.9</td>
</tr>
<tr>
<td>Automated Device %</td>
<td>71.0</td>
</tr>
<tr>
<td>Bayesian Spam Detection</td>
<td>69.5</td>
</tr>
<tr>
<td>Manual Device %</td>
<td>69.2</td>
</tr>
<tr>
<td>Registration Date</td>
<td>62.9</td>
</tr>
<tr>
<td>Mention Ratio</td>
<td>56.2</td>
</tr>
<tr>
<td>Link Safety</td>
<td>49.3</td>
</tr>
<tr>
<td>Hashtag Ratio</td>
<td>47.0</td>
</tr>
<tr>
<td>Followers to Friends Ratio</td>
<td>45.3</td>
</tr>
<tr>
<td>Account Verification</td>
<td>35.0</td>
</tr>
</tbody>
</table>

No single tree works well, a forest gives the best accuracy
Decision Maker

<table>
<thead>
<tr>
<th>Actual</th>
<th>Classified</th>
<th>Total</th>
<th>True Pos</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human</td>
<td>Cyborg</td>
<td>Bot</td>
</tr>
<tr>
<td>Human</td>
<td>1972</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>Cyborg</td>
<td>65</td>
<td>1833</td>
<td>102</td>
</tr>
<tr>
<td>Bot</td>
<td>2</td>
<td>46</td>
<td>1952</td>
</tr>
<tr>
<td>Avg</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Minimal misclassification between human and bot
Training dataset (spam/nonspm) missing inputs:
• Bot tweets with no URLs left out
• Human and cyborg tweets with URLs left out
Who are the users?

Registered users: ~1 billion

Idle users: ~700 millions

Active users: ~300 millions
Twitter composition

Only one half is human!!!
CASA Paradigm

Social scripts are the same in both human-human and human-computer interaction
CASA Paradigm

Is the *communication quality* perceived differently between a bot and a human?
Communication Quality
Communication Quality

- Source Credibility
- Interpersonal Attraction
- Communication Competence
- Intent to Interact
Testing Perception

Bots helping people...

Researchers helping people...
Communication Quality - Results

Twitterbot got generally lower scores than the human, but it is still perceived as credible source.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Twitter bot</th>
<th></th>
<th>Twitter human</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td></td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td>5.12_a (1.07)</td>
<td></td>
<td>5.31_a (1.05)</td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>5.01_a (1.13)</td>
<td></td>
<td>5.27_a (1.03)</td>
<td></td>
</tr>
<tr>
<td>Caring</td>
<td>4.72_a (1.26)</td>
<td></td>
<td>4.95_a (1.05)</td>
<td></td>
</tr>
<tr>
<td>Task attraction</td>
<td>3.40_a (0.63)</td>
<td></td>
<td>3.58_b (0.62)</td>
<td></td>
</tr>
<tr>
<td>Social Attraction</td>
<td>2.61_a (0.74)</td>
<td></td>
<td>2.87_b (0.63)</td>
<td></td>
</tr>
<tr>
<td>Communication competence</td>
<td>2.98_a (0.72)</td>
<td></td>
<td>3.09_a (0.73)</td>
<td></td>
</tr>
<tr>
<td>Intent to interact</td>
<td>2.62_a (1.08)</td>
<td></td>
<td>2.55_a (1.00)</td>
<td></td>
</tr>
</tbody>
</table>
Communication Quality - Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Twitter bot</th>
<th></th>
<th>Twitter human</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>(SD)</td>
<td>M</td>
<td>(SD)</td>
</tr>
<tr>
<td>Competence</td>
<td>5.12_a</td>
<td>(1.07)</td>
<td>5.31_a</td>
<td>(1.05)</td>
</tr>
<tr>
<td>Character</td>
<td>5.01_a</td>
<td>(1.13)</td>
<td>5.27_a</td>
<td>(1.03)</td>
</tr>
<tr>
<td>Caring</td>
<td>4.72_a</td>
<td>(1.26)</td>
<td>4.95_a</td>
<td>(1.05)</td>
</tr>
<tr>
<td>Task attraction</td>
<td>3.40_a</td>
<td>(0.63)</td>
<td>3.58_b</td>
<td>(0.62)</td>
</tr>
<tr>
<td>Social attraction</td>
<td>2.61_a</td>
<td>(0.74)</td>
<td>2.87_b</td>
<td>(0.63)</td>
</tr>
<tr>
<td>Communication competence</td>
<td>2.98_a</td>
<td>(0.72)</td>
<td>3.09_a</td>
<td>(0.73)</td>
</tr>
<tr>
<td>Intent to interact</td>
<td>2.62_a</td>
<td>(1.08)</td>
<td>2.55_a</td>
<td>(1.00)</td>
</tr>
</tbody>
</table>

What if they rated the source and not to the author?
Communication Quality - Results

Followers can perceive twitterbots as **credible**, **attractive**, **competent in communication**

and can have **intention to interact**
Summary

- 53%
- 36%
- 11%
Summary
Summary
Thank you