



## Machine Learning Driven Social Engineering

**Charles Herring**

Co-Founder, CTO

me@CharlesHerring.com

<https://CharlesHerring.com>

@charlesherring

# About Charles



**1995-2002:** Forward Deployed US Navy Hornet Avionics Tech

**2002-2005** US Naval Postgraduate School Network Security  
Group Division Officer  
*sk3wl Of r00t* team member



**2003-2008:** InfoWorld Test Center  
Contributing Product Reviewer – Network and Information  
Security

**2005-2012** DoD Security, Data & Workflow Consultant



**2012-2016:** Consulting Security Architect for Lancope then  
Cisco Systems

**2016** CTO & co-founder at WitFoo



# Disclaimer

*Information in this presentation is intended to protect against illegal or harmful social engineering and human manipulation or to be used in benevolent endeavors.*

**Please do good and obey the law.**



# Old School Social Engineering

- Based on “Confidence Games” (Cons)
- Target a “mark” to achieve inappropriate action
- Requires understanding of “mark” (agent)
- Manipulates existing human vulnerability



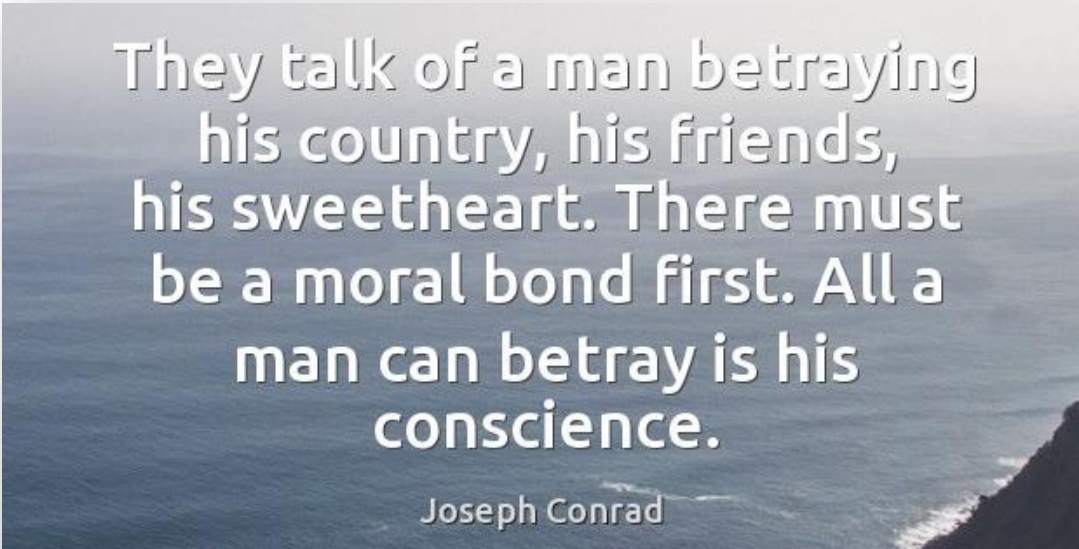
# Second Generation Social Engineering

- Bulk, mass attacks
- Phishing
- Robocalls



# Social Engineering with ML

- Need the Cohort (not a specific agent) to take action
- Long term training will transform agent values (create vulnerability)

A quote by Joseph Conrad is displayed in white text with a subtle drop shadow over a background image of a coastline. The image shows a dark, rocky shore in the foreground, with the ocean extending to the horizon under a hazy, overcast sky. The text is centered and reads: "They talk of a man betraying his country, his friends, his sweetheart. There must be a moral bond first. All a man can betray is his conscience."

They talk of a man betraying  
his country, his friends,  
his sweetheart. There must  
be a moral bond first. All a  
man can betray is his  
conscience.

Joseph Conrad



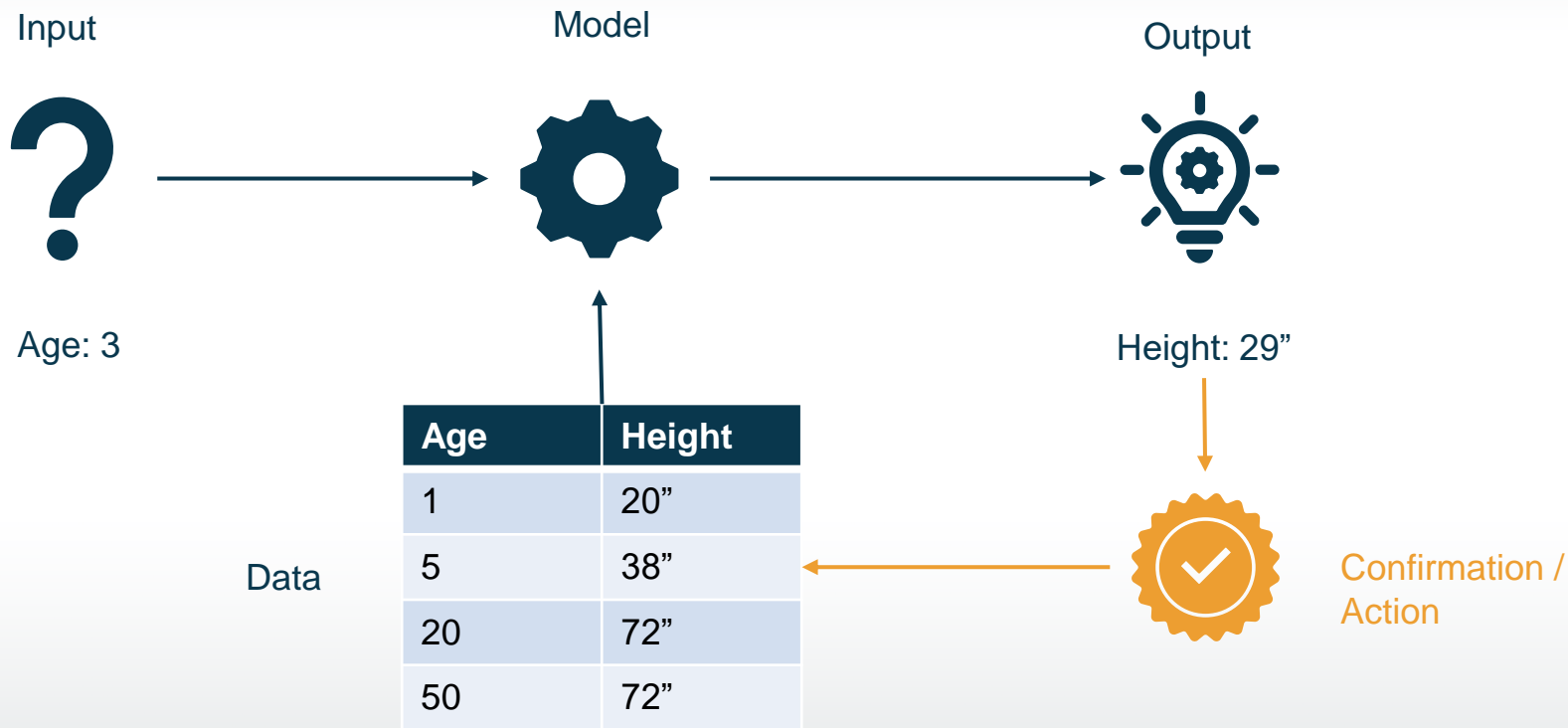
# Relentless Manipulation

- *“We’ve created a world in which online connection has become primary. Especially for younger generations. And yet, in that world, anytime two people connect, the only way it’s financed is through a sneaky third person who’s paying to manipulate those two people. So, we’ve created an entire global generation of people who were raised within a context with the very meaning of communication, the very meaning of culture, is manipulation.” — Jaron Lainer in *The Social Dilemma**

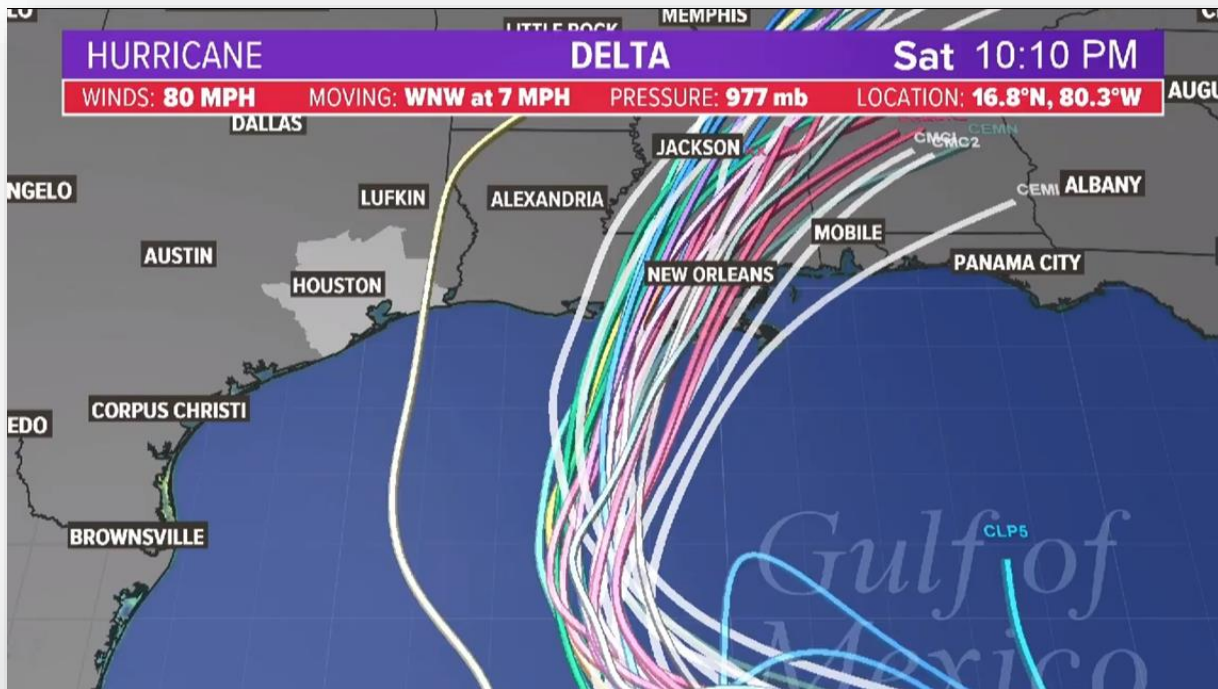
# *Machine Learning Basics*



# Supervised Learning



# Data and Models



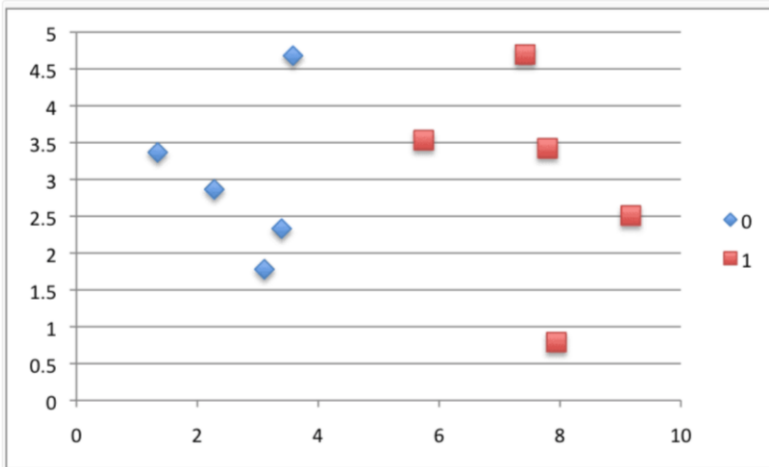
# Common Models

- *Naive Bayes*
- Nearest Neighbor
- Decision Trees
- Linear Regression
- Support Vector Machines (SVM)
- Neural Networks

<https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861>

# Naïve Bayesian Python

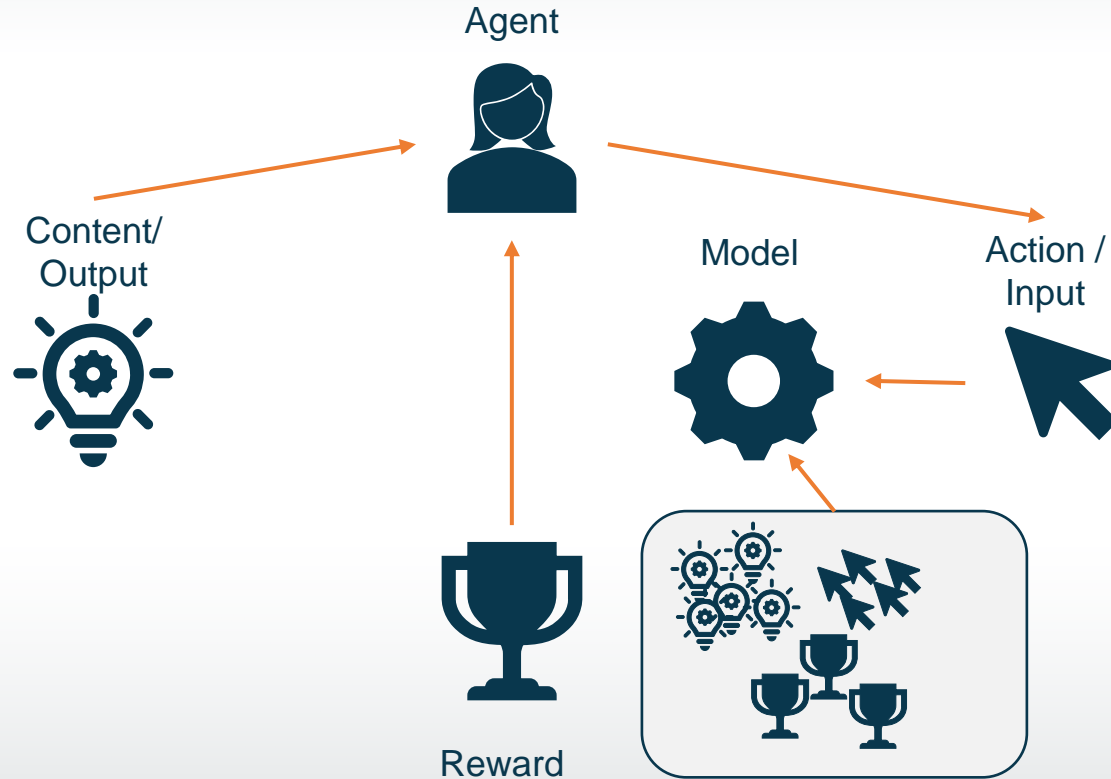
- <https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/>



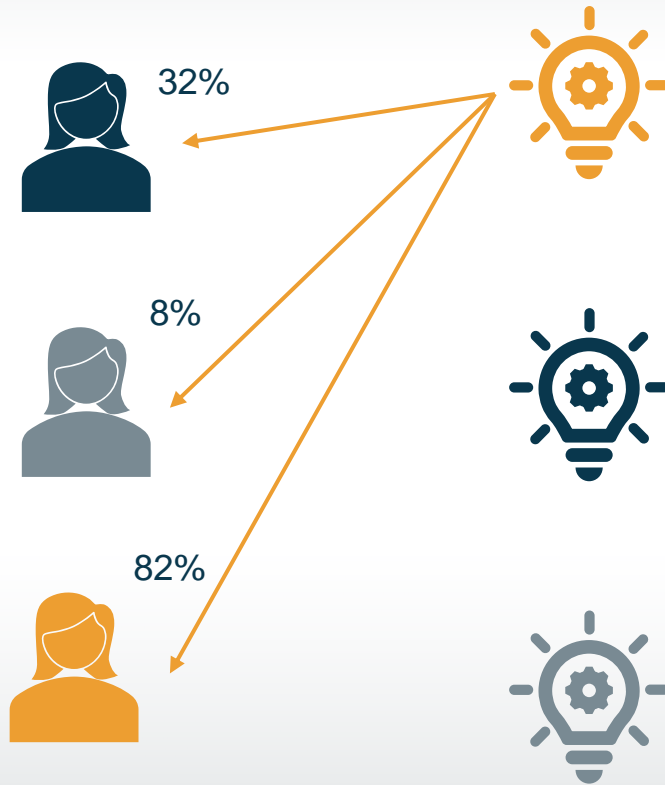
Scatter Plot of Small Contrived Dataset for Testing the Naive Bayes Algorithm

```
1 # Example of separating data by class value
2
3 # Split the dataset by class values, returns a dictionary
4 def separate_by_class(dataset):
5     separated = dict()
6     for i in range(len(dataset)):
7         vector = dataset[i]
8         class_value = vector[-1]
9         if (class_value not in separated):
10             separated[class_value] = list()
11         separated[class_value].append(vector)
12     return separated
13
14 # Test separating data by class
15 dataset = [[3.393533211, 2.331273381, 0],
16            [3.110073483, 1.781539638, 0],
17            [1.343808831, 3.368360954, 0],
18            [3.582294042, 4.67917911, 0],
19            [2.280362439, 2.866990263, 0],
20            [7.423436942, 4.696522875, 1],
21            [5.745051997, 3.533989803, 1],
22            [9.172168622, 2.511101045, 1],
23            [7.792783481, 3.424088941, 1],
24            [7.939820817, 0.791637231, 1]]
25 separated = separate_by_class(dataset)
26 for label in separated:
27     print(label)
28     for row in separated[label]:
29         print(row)
```

# Reinforcement Learning



# A/B Testing of Output / Signal / Content



# Types of Content

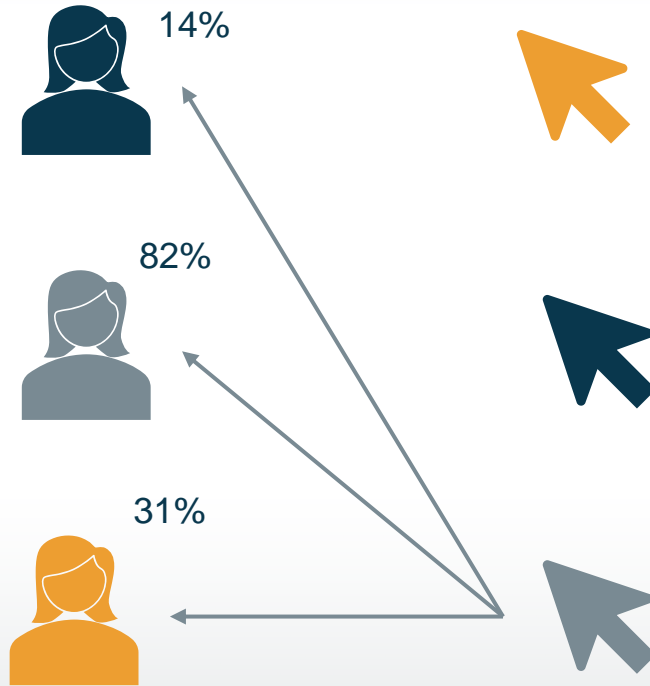
- Text Ad
- Video Ad
- Promoted Pages
- Email
- Snail Mail
- Images



# Message Considerations

- Alignment with Values
- Suspicion Triggers
- Language & Dialect Use
- Use of Nostalgia
- Trauma Triggers (Crossing a line)
- Bias Reinforcement
- Trusted Corroboration
- Social Corroboration
- Pain vs. Reward

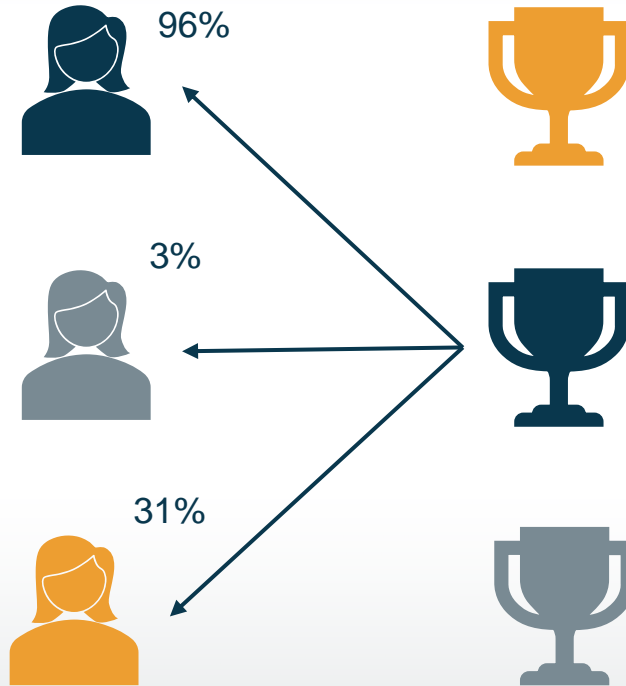
# A/B Testing of Action



# Types of Actions

- Click to Page
- Watch Video
- Play Game
- Fill out Form
- Join a Group
- Install an Application
- Comment / Engage

# A/B Testing of Reward



# Types of Reward

- Public Praise – like, retweet, comment
- Artificial Prize - badge, points
- Physical Prize - cash, gift card, prize, beer, meal
- Education Credit – CEU, certification
- Celebrity Interaction – virtual or physical
- Promise of job
- Romantic Relationship or Encounter
- Confirmation Reinforcement

# *Cohort Targeting*

# Common Cohort Dimensions

- Gender
- Age
- Ethnicity
- Religious Belief
- Income
- Location
- Family Status
- Job Type
- Employer Type
- Education Level
- Affiliations
- Relationships
- Entertainment
- Credit Rating
- Net Worth
- Political Association



# Facebook Audiences

Create Audience

Locations

United States

United States

Search Locations

Detailed Targeting

Include people who match

Behaviors > Politics (US)

Likely engagement with US political content (conservative)

Demographics > Education > Education Level

College grad

Interests > Additional Interests

Grand Rapids, Michigan

The Hacking Universe

Interests > Entertainment > Music

Add demographics, interests or behaviors

Suggestions Browse

Your audience selection is **fairly broad**.

Potential Audience Size: 140,000,000 people

Cancel Save

Income

Household income: top 10% of ZIP codes (US)

Household income: top 10%-25% of ZIP codes (US)

Household income: top 25%-50% of ZIP codes (US)

Household income: top 5% of ZIP codes (US)

Work

Employers

Industries

Job Titles

Soccer

Friends of Soccer fans

Soccer fans (high content engagement)

Soccer fans (moderate content engagement)

# LinkedIn Audience

## Who is your target audience?

Include people who have **ANY** of the following attributes:



### Member Interests

Open Source Software

**AND** also have **ANY** of the following attributes:



### Job Titles (Current)

Software Engineer

**AND** also have **ANY** of the following attributes:



### Company (Current Jobs)

Steelcase

**Narrow** audience further

**Exclude** people by audience attributes and Matched Audiences

LinkedIn tools may not be used to discriminate based on personal characteristics like gender, age, or actual or perceived race/ethnicity. [Learn more](#)

[Learn more about targeting criteria](#)

<b>Audiences</b> Use your data to retarget website visitors or reach known contacts and accounts	Company Demographics Education Job Experience Interests and Traits
<b>Audience attributes</b> Add targeting criteria like job title, industry, or skills	

Useable on Bing Search

# Google Audience

**People:** who you want to reach  
Define your **Audiences**, **Demographic**, or both

## Demographics

Select your demographic targeting ⓘ

Gender	Age	Parental status	Household Income
<input checked="" type="checkbox"/> Female	<input checked="" type="checkbox"/> 18 - 24	<input checked="" type="checkbox"/> Not a parent	<input checked="" type="checkbox"/> Top 10%
<input checked="" type="checkbox"/> Male	<input checked="" type="checkbox"/> 25 - 34	<input checked="" type="checkbox"/> Parent	<input checked="" type="checkbox"/> 11 - 20%
<input checked="" type="checkbox"/> Unknown ⓘ	<input checked="" type="checkbox"/> 35 - 44	<input checked="" type="checkbox"/> Unknown ⓘ	<input checked="" type="checkbox"/> 21 - 30%
	<input checked="" type="checkbox"/> 45 - 54		<input checked="" type="checkbox"/> 31 - 40%
	<input checked="" type="checkbox"/> 55 - 64		<input checked="" type="checkbox"/> 41 - 50%
	<input checked="" type="checkbox"/> 65+		<input checked="" type="checkbox"/> Lower 50%
	<input checked="" type="checkbox"/> Unknown ⓘ		<input checked="" type="checkbox"/> Unknown ⓘ

⚠ Note: Household income targeting is only available in select countries. [Learn more](#)

## Audiences

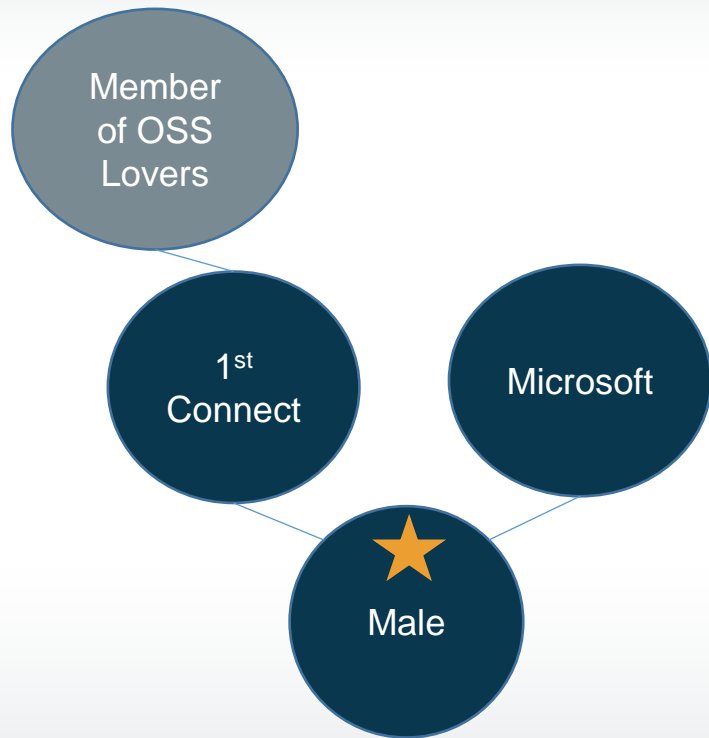
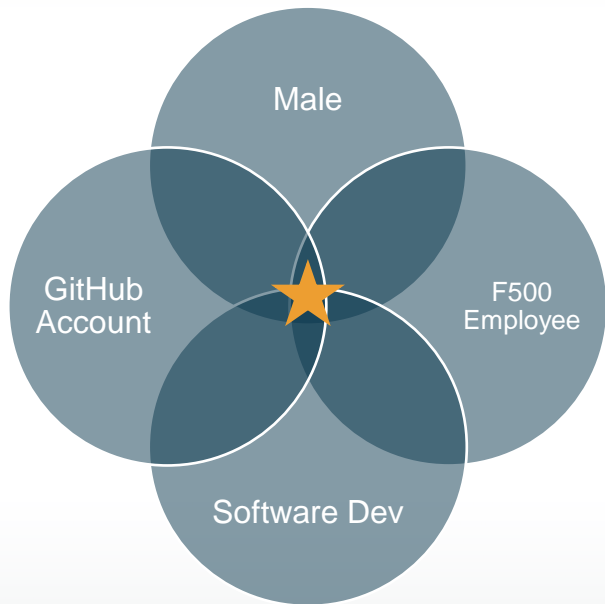
Select audiences to define who should see your ads. You can create new audiences in [Audience Manager](#). ⓘ

SEARCH	BROWSE	7 selected	CLEAR ALL
Who they are (Detailed demographics)	>	Detailed demographics	
What their interests and habits are (Affinity)	>	Employment > Company Size Very Large Employer (10k+ Employees)	✕
What they are actively researching or planning (In-market and life events)	>	Employment > Industry Manufacturing Industry	✕
How they have interacted with your business (Remarketing and similar audiences)	>	Parental Status > Parents Parents of Preschoolers (4-5 years)	✕
Your combined audiences (Combined audiences)	>	Marital Status Single	✕
Your custom audiences (Custom audiences)	>	Homeownership Status Homeowners	✕
		Education > Highest Level of Educational Attainment Bachelor's Degree	✕
		Affinity audiences	

# Other Sources of Information

- Breach Data (Cambridge Analytica, etc)
- Intellius
- Scraping Social Media Sites
- Public Records

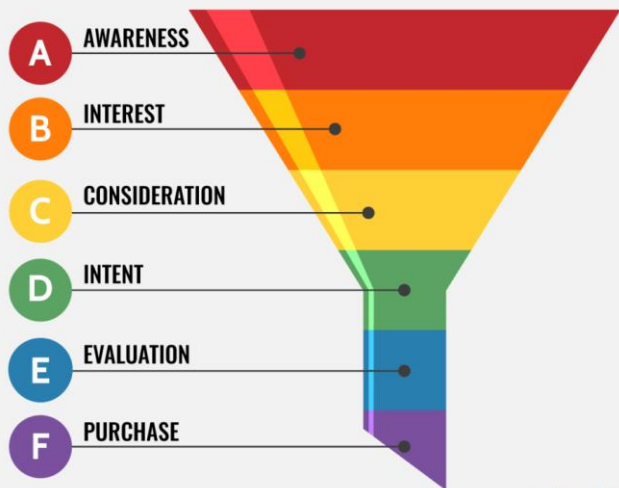
# Cohort Collision vs Relationship



# *Launching Campaigns*

# Humans Move Slowly

## THE TRADITIONAL SALES FUNNEL



@2021 DK New Media, LLC. All Rights Reserved.

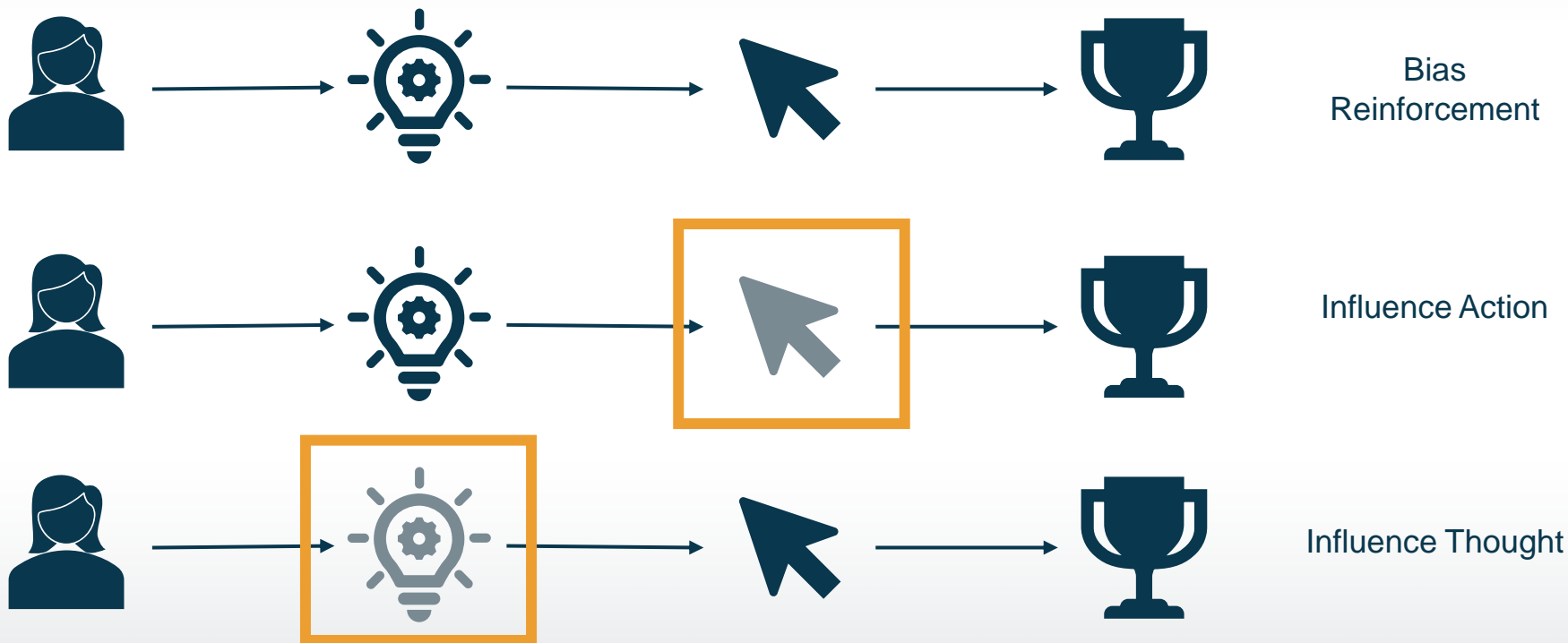


Sales cycles as detailed here can take months or years to complete as agents progress through the process.

Successful progressions will have small, incremental steps requiring the lowest amount of agent change per stage.



# Agent Transformation



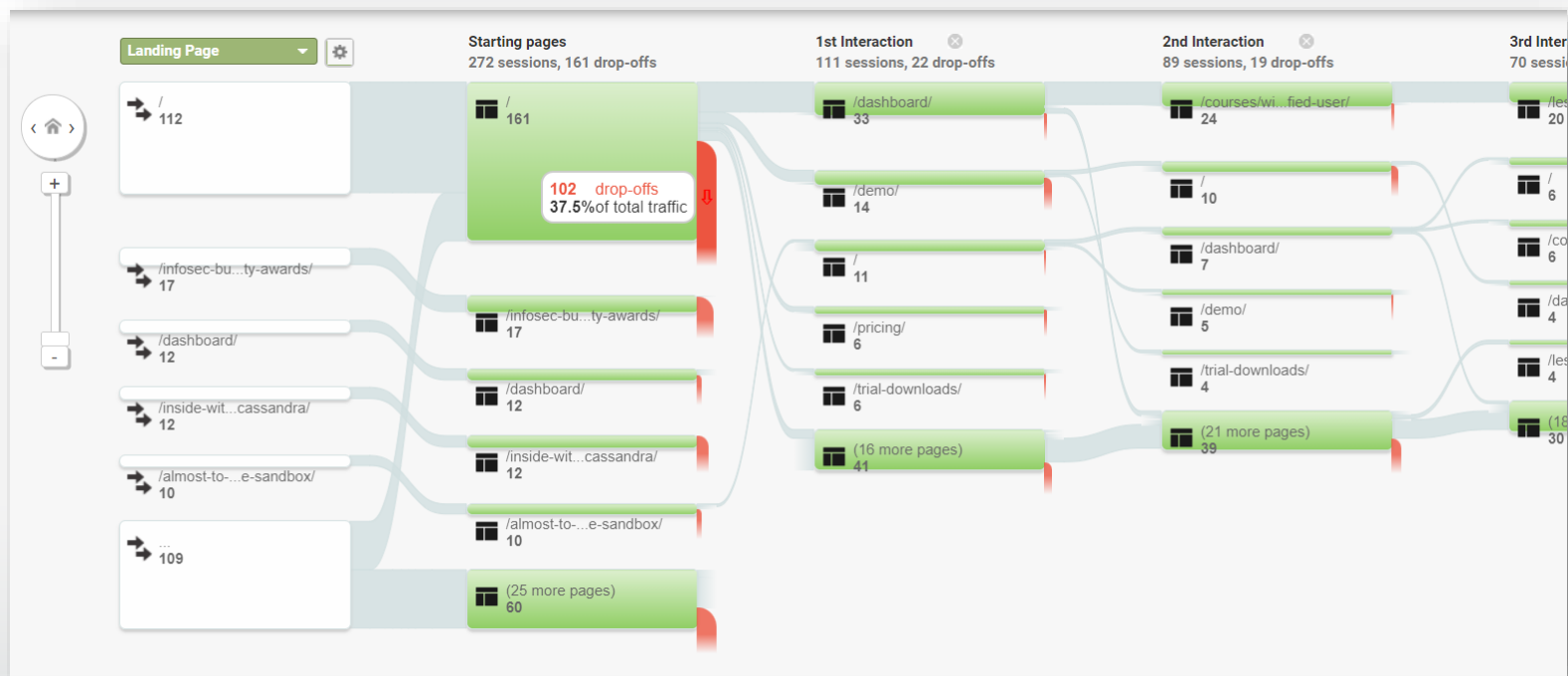
# Key Tracking Metrics

- Agent ID (Preferred) or Cohort ID
- Content ID
- Action ID
- Action State
- Proposed Reward ID

ID	Cohort	CID	AID	AS	RID
1000	1	3	2	0	1
1002	2	3	2	1	1
1000	1	4	2	1	2
1001	2	3	2	1	1

[https://test.com/page.php?agent\\_id=1001&content\\_id=3&action\\_id=2&state=fail&reward\\_id=4](https://test.com/page.php?agent_id=1001&content_id=3&action_id=2&state=fail&reward_id=4)

# Behavior Flow Analysis

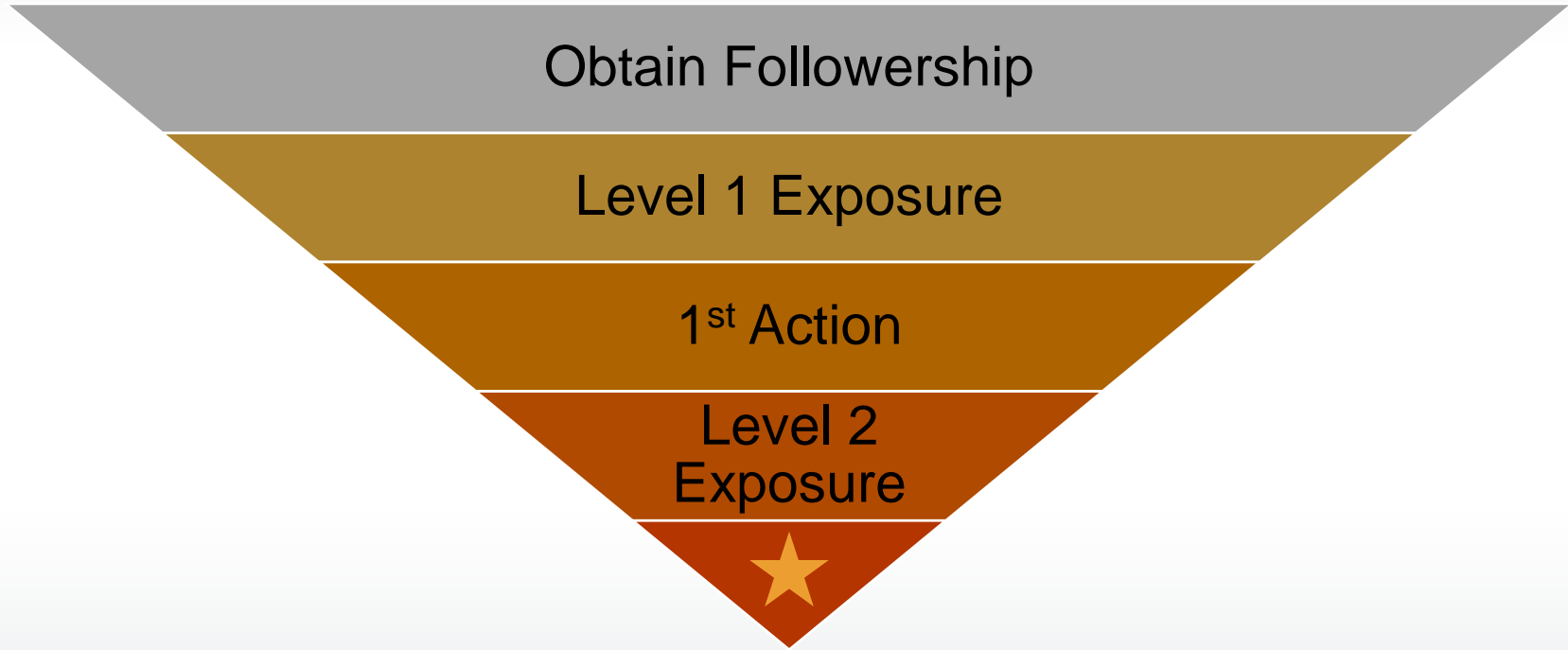


# Improving Cohort Definitions

- What characteristics are shared among funnel exits?
- Are those characteristics missing in continuations?
- Tracking interaction counts to conversion

ID	Age	Politics	Income	Role	Music
1003	57	Near Left	\$120k	PM	EDM
1198	45	Far Left	\$200k	Exec	EDM
2876	33	Near Right	\$85k	Dev	EDM

# Multi-stage Cultivation



# Maintaining a Campaign

- Data Mining for Cohort Definition
- Large, diverse, effective content
- Ability to maintain audience (Groups, followers, mail lists, games)
- A/B testing of action delivery
- Reward types
- Decent Model (Naive Bayes)
- Growing Dataset for Model
- Detection Avoidance





## Machine Learning Driven Social Engineering

**Charles Herring**

Co-Founder, CTO

me@CharlesHerring.com

<https://CharlesHerring.com>

@charlesherring