“SELF-PROGRAMMABLE” AGENTS
KEEPING UP BOT FOLLOWERS AND FOLLOWINGS

Elizabeth A. Ondula
Ph.D student Computer Science

VITERBI SCHOOL OF ENGINEERING

Summer Research 2020, STEM BYTES TALK
My Research Interest

To address

People Edge and IoT Systems

Social Infrastructure Challenges

Design Thinking

My research

Data Input

 RL AGENTS

Output: informs

AI TOOLS DESIGN AND DEVELOPMENT

Teach

Teach
TOPIC: SOCIAL ENGINEERING BY AI: GOOD OR BAD FOR SECURITY?
What is Social Engineering Attacks?

The use of human interaction skills to gain access to computing systems

Technology used today: Deepfakes

Figure 1. The process of creating a real-time imitation of a human neural network in DeepFake technology

Image Reference: Y Shtefaniuk et al
Forms of Social Engineering attacks

Phising – soliciting personal information such as email. Mostly pose as trustworthy orgs e.g USC

Vishing – leverages voice communication.

Smishing – text/SMS containing malicious links e.g to porn sites that are notorious for malware

Definitions: https://us-cert.cisa.gov/ncas/tips/ST04-014
IMPACT ON CONTEMPORARY SOCIETIES

Elections    Education    Food Systems
Finding Housing

Trade Systems
RESEARCH QUESTION

Given the number of disappointed populations globally experiencing both environmental and economic losses due to energy consumption by automated software agents, how can deepfakes today be used to identify trolls or bots on social networking sites?
EXEMPLARY SCENARIO

Who are my ‘real’ friends that I can rely on for personal matters?

SOCIAL 911 😊

A friend in this context is considered as a node stored on a graph DB with 2 primary edges:

- Follower
- Following

To a friend, doesn’t matter the label, just sharing memories and being.
HYPOTHESIS

Trolls or bots are accounts that immediately respond once a post has been published.

Troll or bot is real when it responds with a emoji sequence that is meaningful.

Troll or bot is fake when it responds in a jibberish manner.
DISTRIBUTED SYSTEM DESIGN AND FLOW

- Facebook UI (Black)
- Facebook UI (White)
- Data Pulling Service
- WhatsApp Monitor
- Troll Detector Agent
- Instagram Counter

INPUT CHANNEL
Off-Cloud
Home Network

SERVICES CHANNEL
On-Cloud
Software-Defined Network

CLOUD INFRASTRUCTURE

LIKES, LOVES, LAUGHTER
From existing contact

OUTCOME
Mobile Network (3G/4G/5G LTE)
Definition

A **programmable RL agent** is one that not only computes its own beliefs by logically reasoning through its own knowledgebase but also rewards itself which in turn rewards other programmable software agents interconnected to it.

The goal of the agent is simply to flag bots/trolls
Algorithm.

Input: GET (FOLLOWERS, FOLLOWING) from (SN API)
Method:

• Initialize with random
• Initialize q-learning agent
• Wait (time)
• Try (likes OR loves, AND comments to new faces):
  • Catch (Phising bots with fewer than 10 and < 10 min
    • Following response)

Expected Output: POST (Similar interests and
*shares)
SUMMARY: RESEARCH QUESTION

Given the number of disappointed populations globally experiencing both environmental and economic losses due to energy consumption by automated software agents, how can deepfakes today be used to identify trolls or bots on social networking sites?
Tool Set

• Instagram account
• Facebook developer
• Mobile device
• Edge device (Raspberry Pi)


Online resources to be updated on my website eondula.github.io/talks/resources
What’s next?

1. Classification of "attack" types based on personality type using
2. Transfer learning to adopt to other domains
3. Open sourcing the project
THANK YOU
QUESTIONS?
FEEDBACK?
ondula@usc.edu