Towards Responsible Human-Agent Collectives

Sarvapali D. (Gopal) Ramchurn

Professor of Artificial Intelligence

Director, Centre for Machine Intelligence

Invited Talk, Thessaloniki, October 2018

Centre For MachineIntelligence

UNIVERSITY OF Southampton

AXA Research Fund

Background

- Professor of AI (2018-) at the University of Southampton
- Director, Centre for Machine Intelligence (2017-)
  - 130+ Researchers
  - Top Group for Impact in the UK
  - 15+ Projects in AI and Machine Learning
- Academic
  - PhD in Multi-Agent Trust and Negotiation (Southampton) – 2004
- Advisory roles:
  - UK Cabinet Office (OI team) (2018-)
  - Tech Startups: Engagetech and UTU Kenya
  - Chief Scientist at North Star Solar (2017-)
- Awards:
  - Best Paper/Nominations AAMAS 2010/11,13,15, IJCAI-JAIR
  - AXA Award for Responsible AI (2018)

Projects

- Past Research Projects:
  - ALADDIN (EPSRC – BAE 2005-2010)
  - IDEAS: Intelligent Decentralised Energy-aware Systems (SECURE LTD 2009-2013)
  - ORCHID: Foundations of Human-Agent Collectives (EPSRC Programme 2011-2016)
  - SEACORES: Fault-Diagnosis on Ships (I-UK 2015-2016)
  - CharIoT: Energy monitoring kit for energy advisors (EPSRC 2015-2017)
  - Human-UAV Teaming (2015-2017)
- Current Projects
  - AXA Responsible AI (2018-2021) £200K
  - Smart Cities and Wearable Tech (EPSRC 2017-2021) £1.2M
  - Autonomous IoT (EPSRC 2016-2019) £800K
  - GCRF BRECCIA (EPSRC 2018-2022) £2.5M

Goals of this Talk

- Explain what human-agent collectives (HACs) are
- Detail some examples of HACs and Applications
- Present some new research directions for Responsible AI/HACs

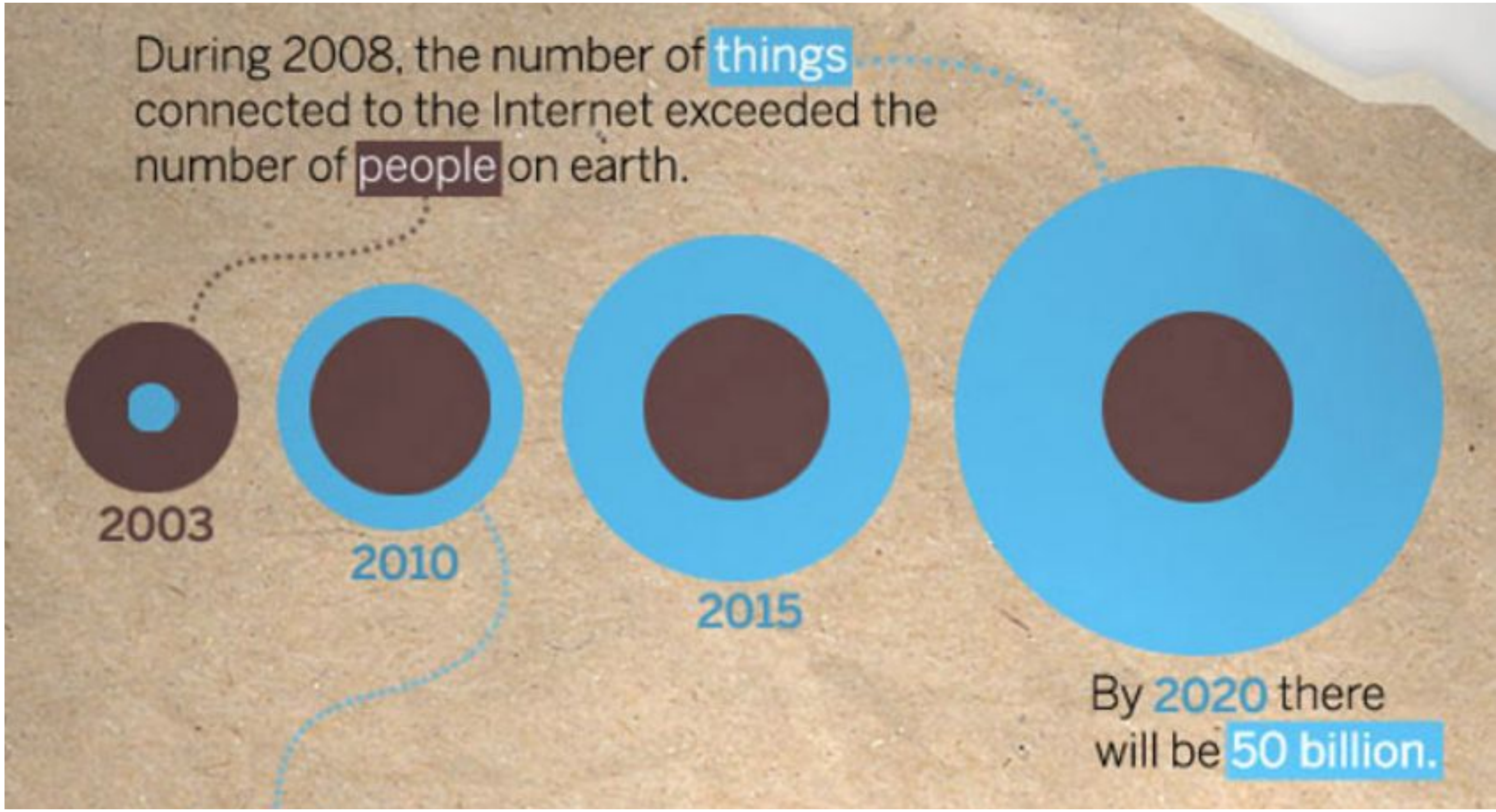
An Era of Information Ubiquity

Connected

Diverse Sources

Measure everything

Mashed up



More helpful computers for a new way of life

(Jennings et al., 2014)

Proactive

Mixed-reality

Machines as collaborators

Human-Agent Collectives (HACs)

Centre For MachineIntelligence

UNIVERSITY OF Southampton

AXA Research Fund

Flexible Autonomy

neither agents, nor humans are always in charge

- Humans act with varying degrees of computer support.
- Agents can act autonomously, other times guided by much closer human involvement.
- Vary depending on context.

Agile Teaming

continually establish and manage collaborative relationships

Groups of agents and humans:

- **Come together** when needed to achieve goals no individual can achieve in isolation
- **Disband** once cooperative action has been successful.

Incentive Engineering

motivate by incentive, rather than diktat

- **Design rewards** so actions that are encouraged generate desirable outcomes.
- Account for human **perception and motivations**

Accountable Information Infrastructure

track information veracity and provenance

- heterogeneous data that has varying degrees of **reliability** and **accuracy**.
- Allow **veracity** and **accuracy** of information to be **confirmed** and **audited**,
  while maintaining **privacy** and **ethics** standards.


The Telegraph

Home Video News World Sport Finance Comment Culture Travel Life Women

Budget | Companies | Comment | Personal Finance | ISAs | Economy | Markets | Property | Ente

Shares | Questor | Market Report | FTSE 100 | Currency | Commodities

HOME » FINANCE » MARKETS

'Bogus' AP tweet about explosion at the White House wipes billions off US markets

The FBI and SEC are to launch investigations after more than £90bn was temporarily wiped off the US stock market when hackers broke into the Twitter account of the Associated Press and announced that two bombs had exploded at the White House, injuring Barack Obama.

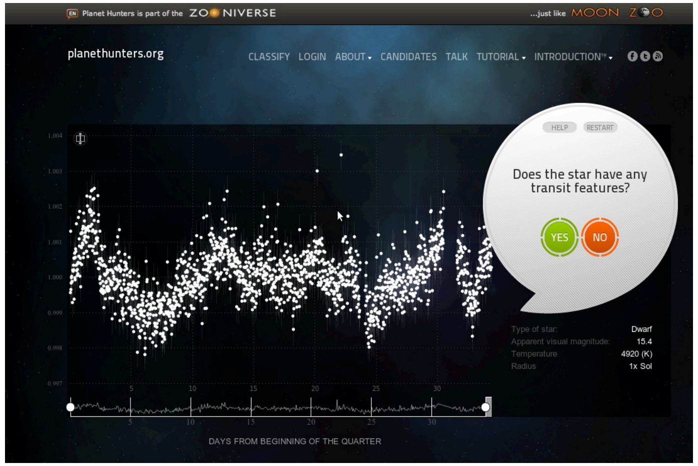
HACs in Smart Energy Systems

Smart Heating Control

Personalised Recommendations & Advice Giving

Electric Vehicle Charging

HACs in Citizen Science



Classifying Galaxies

Hunting for Endangered Species

Centre For MachineIntelligence

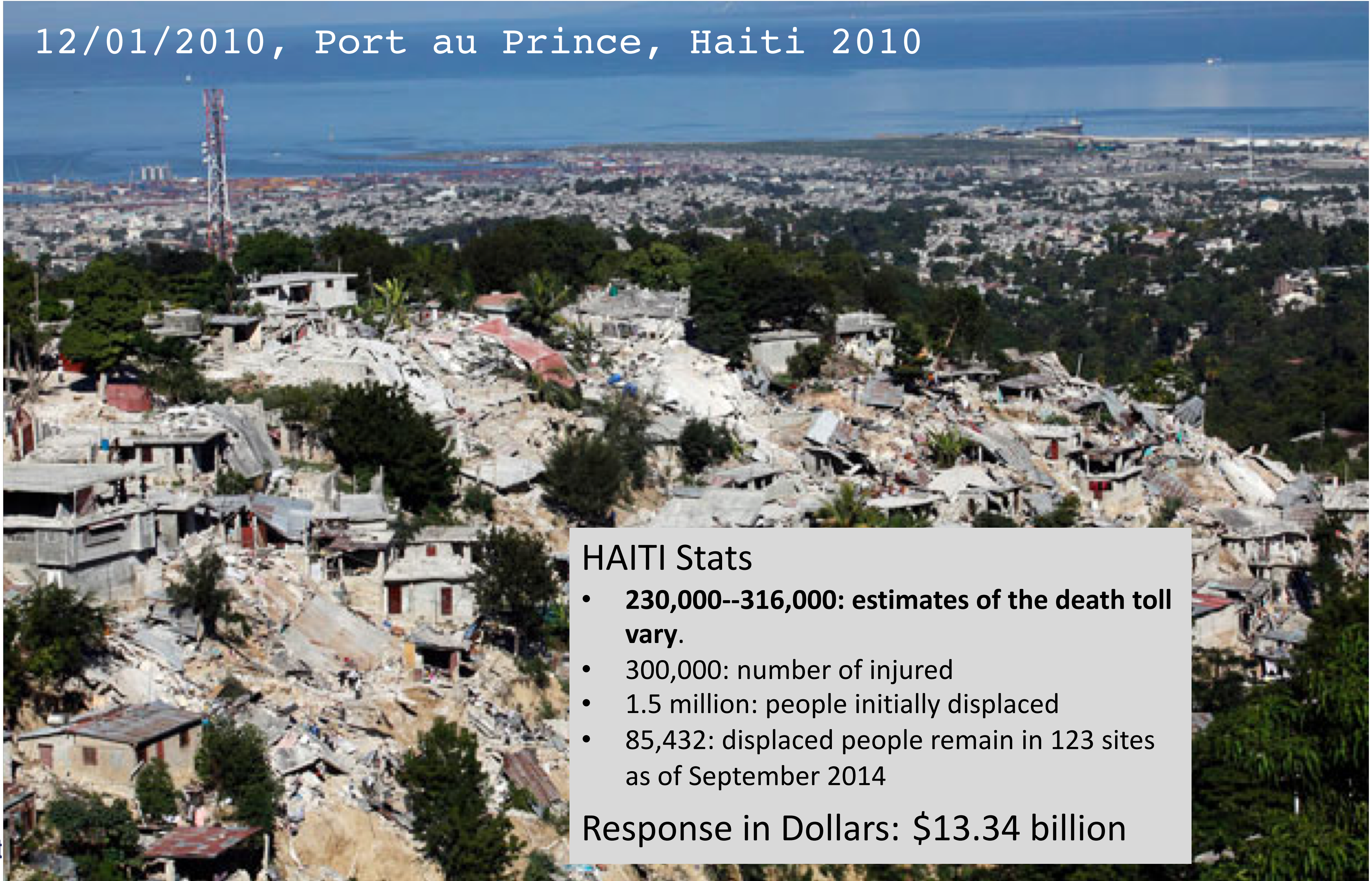
UNIVERSITY OF Southampton

AXA Research Fund

AAMAS Best Paper Applications Track 2015

HACs in Disaster Response

12/01/2010, Port au Prince, Haiti 2010

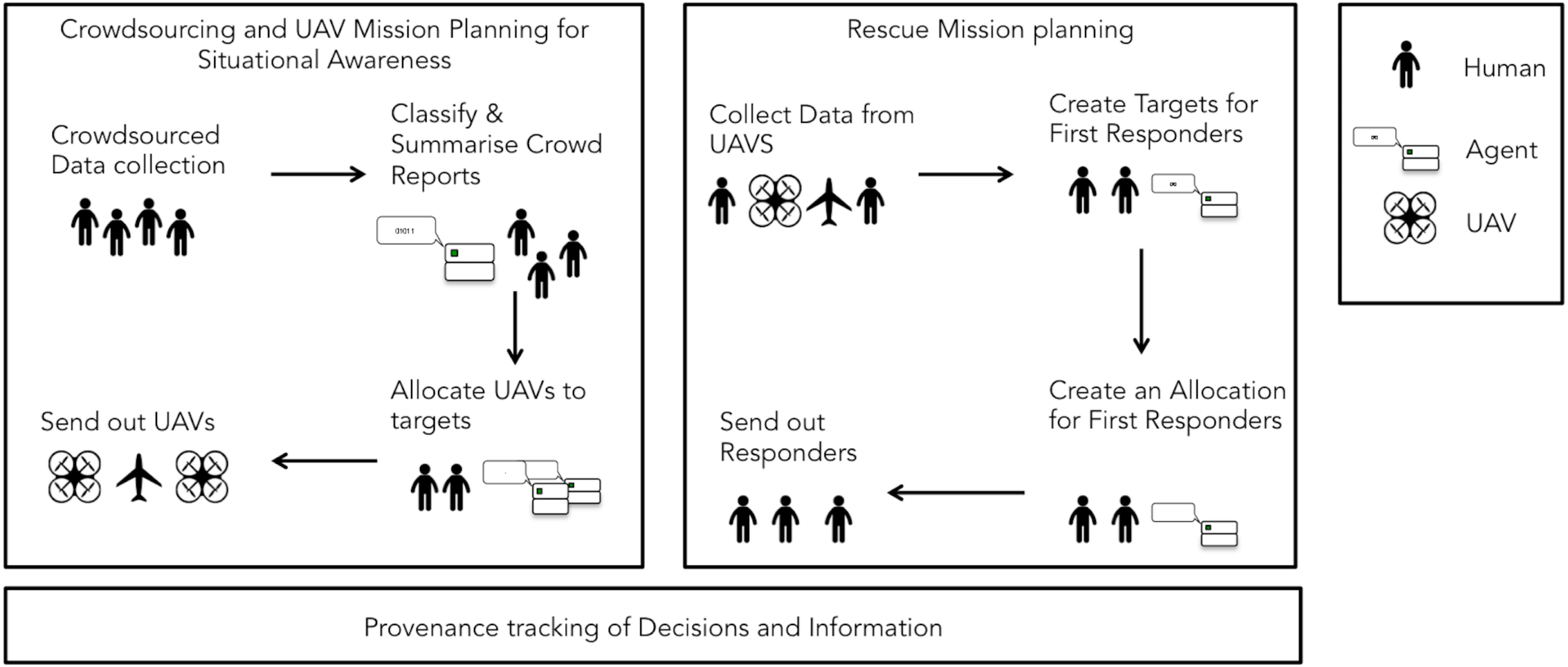


HAITI Stats

- **230,000--316,000: estimates of the death toll vary**.
- 300,000: number of injured
- 1.5 million: people initially displaced
- 85,432: displaced people remain in 123 sites as of September 2014

Response in Dollars: $13.34 billion

A Disaster Response System based on HACs



Centre For MachineIntelligence

UNIVERSITY OF Southampton

AXA Research Fund

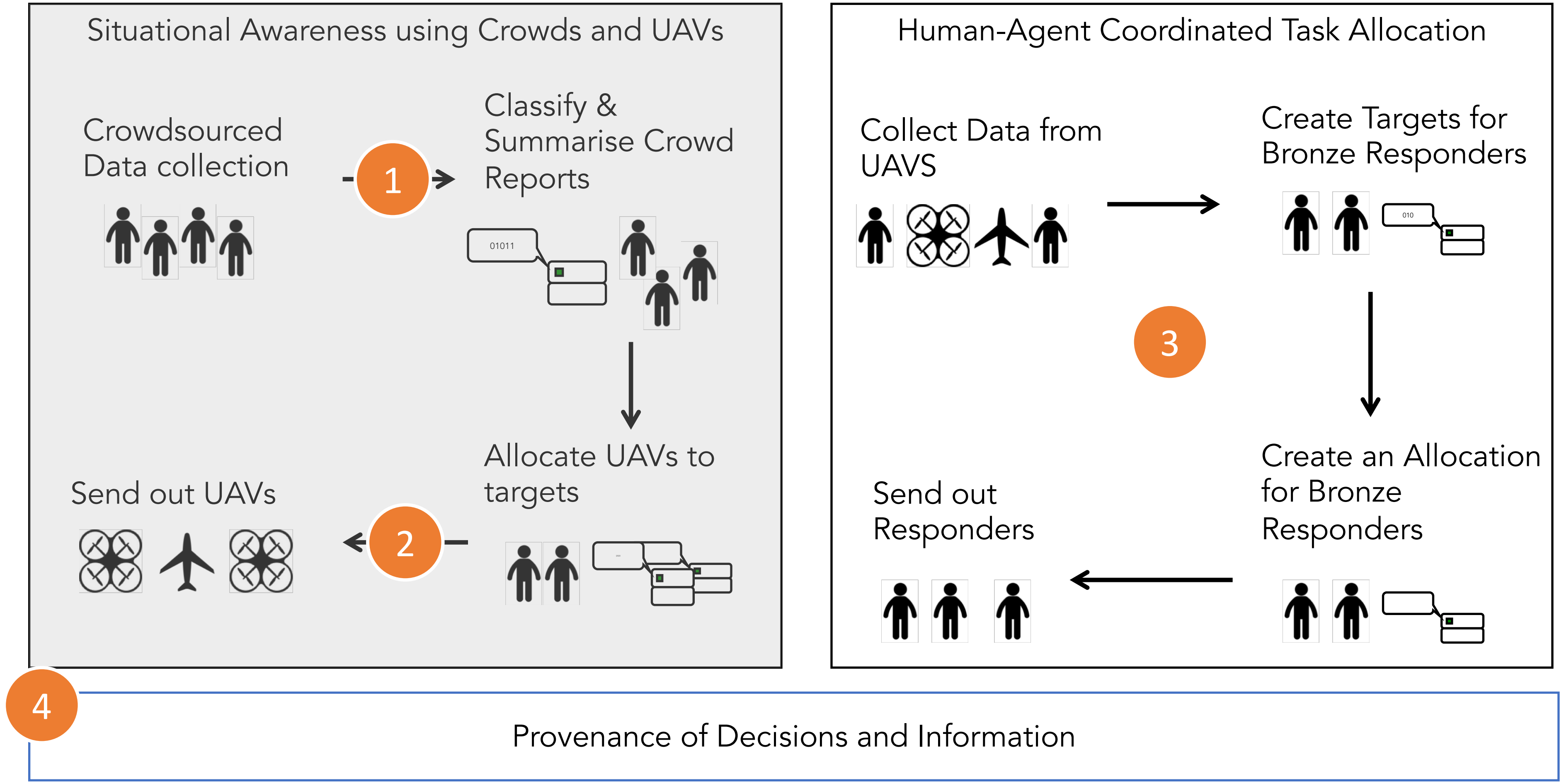
How do we LOCATE casualties and resources?

How to ALLOCATE resources?

How to DEPLOY rescue teams across a large area?

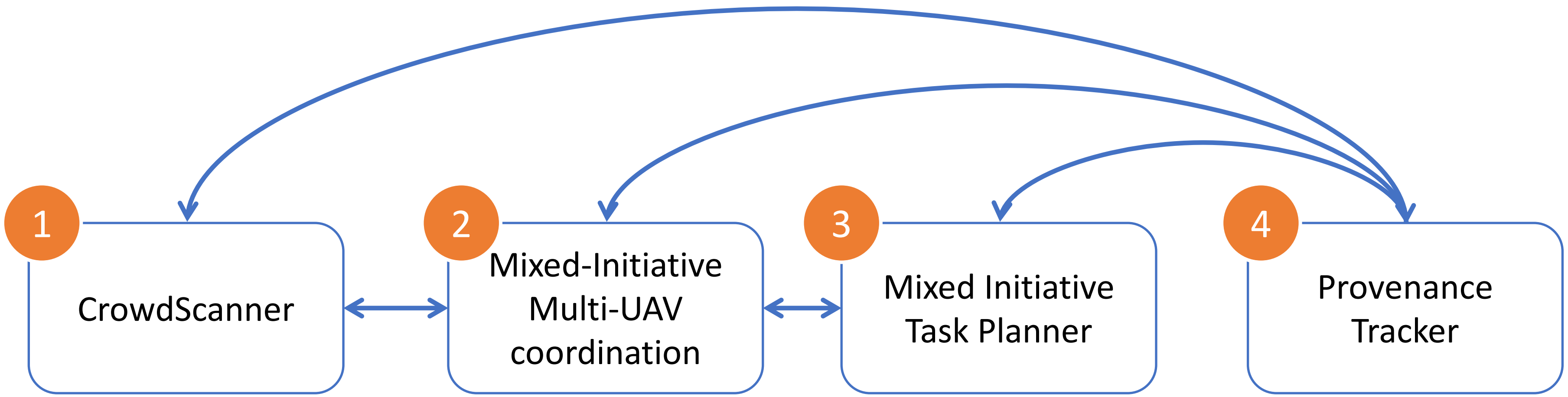
How do we TRUST the information gathered?

Information gathering and coordination loops in HAC-ER



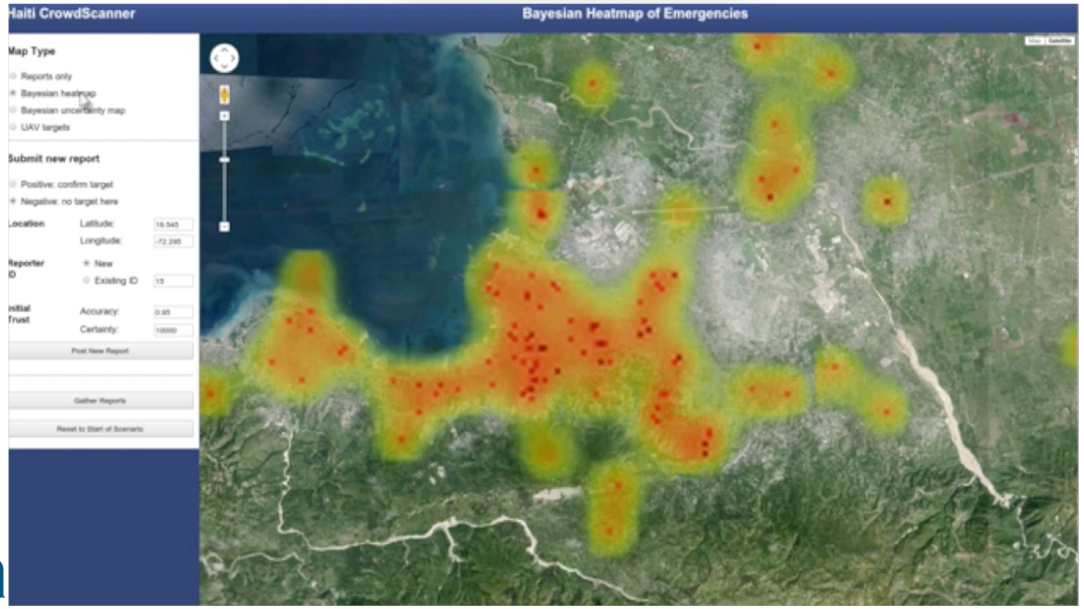
agent human uavs

HAC-ER Modules



CrowdScanner: making sense of crowd reports using human and machine intelligence

- Interpretation
  - Online (imperfect) Crowds + Machine Learning (BCC+ NLP)
  - Hire+Fire algorithm to recruit the best workers and get the best interpretation
- Heatmap creation
  - Gaussian Process to model disaster
  - Fold in trusted reports
  - Use classification output to generate intensity
- Generate targets for UAVs

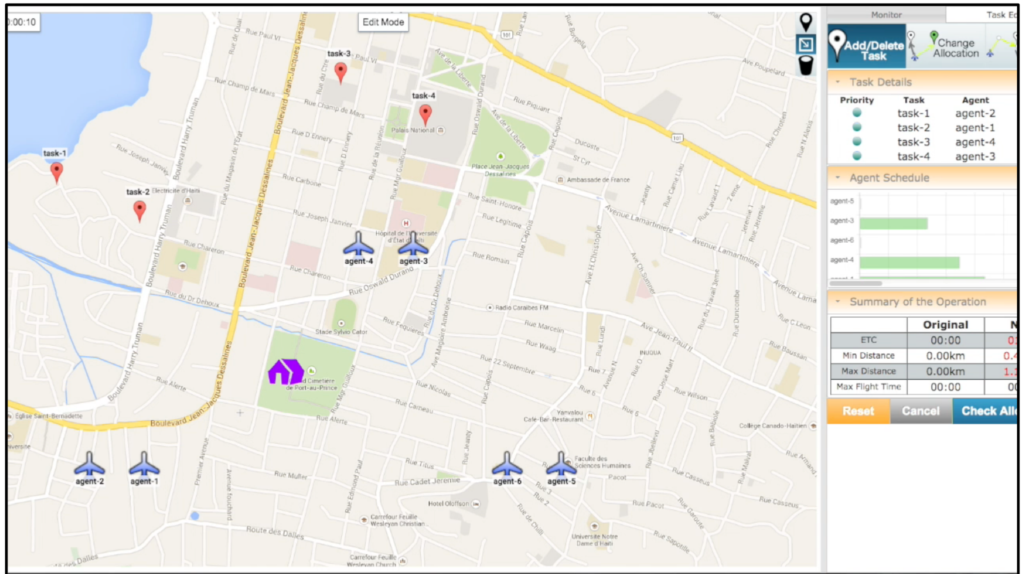


Centre For MachineIntelligence

UNIVERSITY OF Southampton

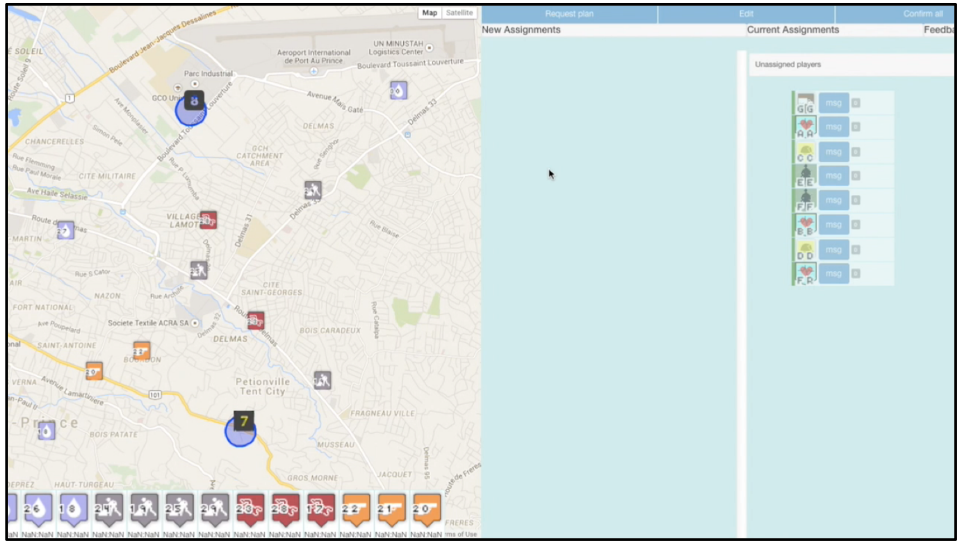
Mixed Initiative Multi-UAV Coordination

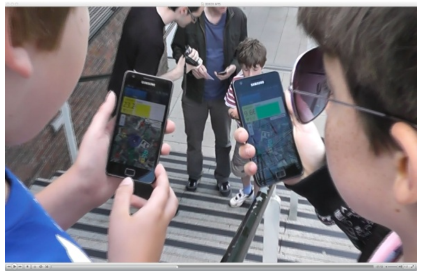
- 1 pilot -> 3+ UAVs
- Heterogeneous UAVs running max-sum
- Flexible Autonomy
  - ‘Adjust’ max-sum plans
  - React to UAV drop-outs
  - Transfer of control between Silver, UAVs, and Bronze operators
- **Validated on real UAVs**
- Tested with 40 users in Lab
- UAVs Targets confirmed for Responders to be deployed



Human-Agent Collaboration for Task Allocation

- Human-Agent Silver team allocate tasks to Bronze responder team
- Agent uses Multi-agent Markov Decision Process
  - Computes best task for each responder, and best path for each task
  - Models environment (buildings and lakes are obstacles)
- Responders get instructions via mobile app
- Trialed in the AtomicOrchid Mixed Reality Game with 100+ users including emergency responders.





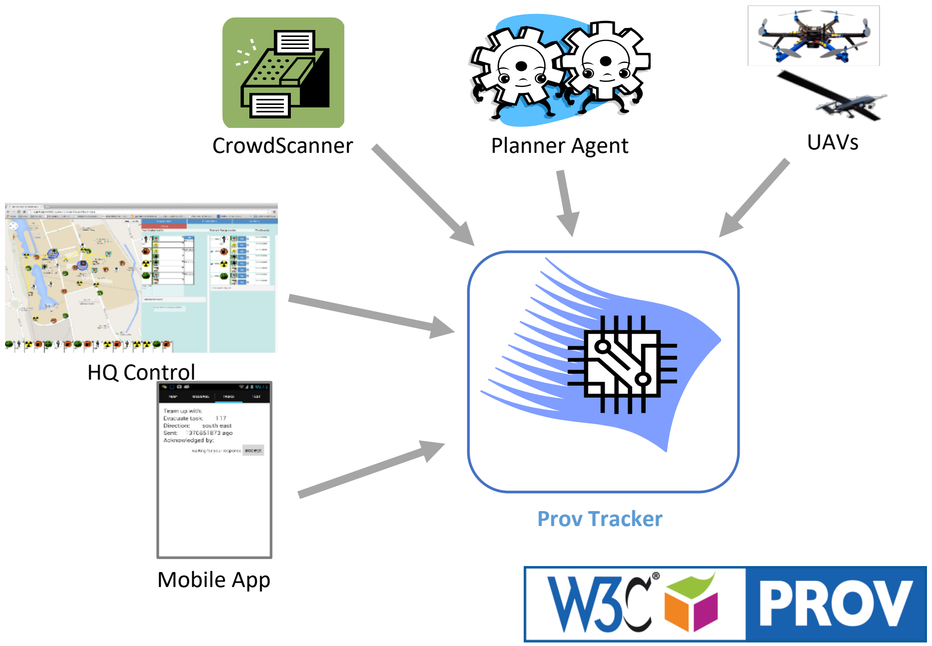
Centre For MachineIntelligence

UNIVERSITY OF Southampton

AXA Research Fund

Supporting Human and Agent Decision Makers using Provenance

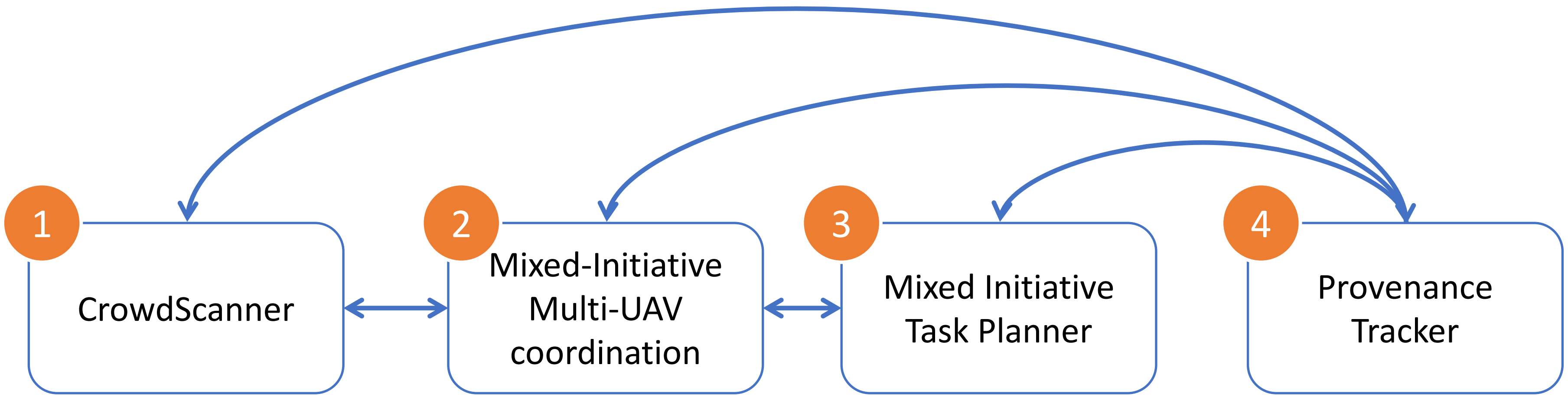
- Timely Decision Support
  - Live monitoring of provenance for changes
  - Ensures the *whole system* reacts to changes
- Post-hoc analysis



Example:

- During the operation, UAVs invalidate targets,
- Prov tracker immediately notifies Silver commanders at HQ
- Prov tracker identifies impacted rescue missions

HAC Interactional Arrangements



Multiple Humans as Sensors
Single Agent
Learning

Multiple Humans Planning
Multiple Agents Planning
Embodied Agents

Multiple Humans Planning
Single Agent Planning

Provenance Agent
Multiple Humans
Multiple Agents

UNIVERSITY OF Southampton

Human-UAV teaming in dynamic and uncertain environments

[dstl]

MBDA MISSILE SYSTEMS

TATA

TATA CONSULTANCY SERVICES

HAC Challenges

- **Organisational** challenges: processes and coordination mechanisms.
- **Interactional** challenges: interfaces and interaction modalities.
- **Accountability** challenges: human and agent as equal partners

Can we define a methodology to design HACs that are Responsible?

What is Responsible AI?

Controllable

Understandable

Trustworthy

Ethical

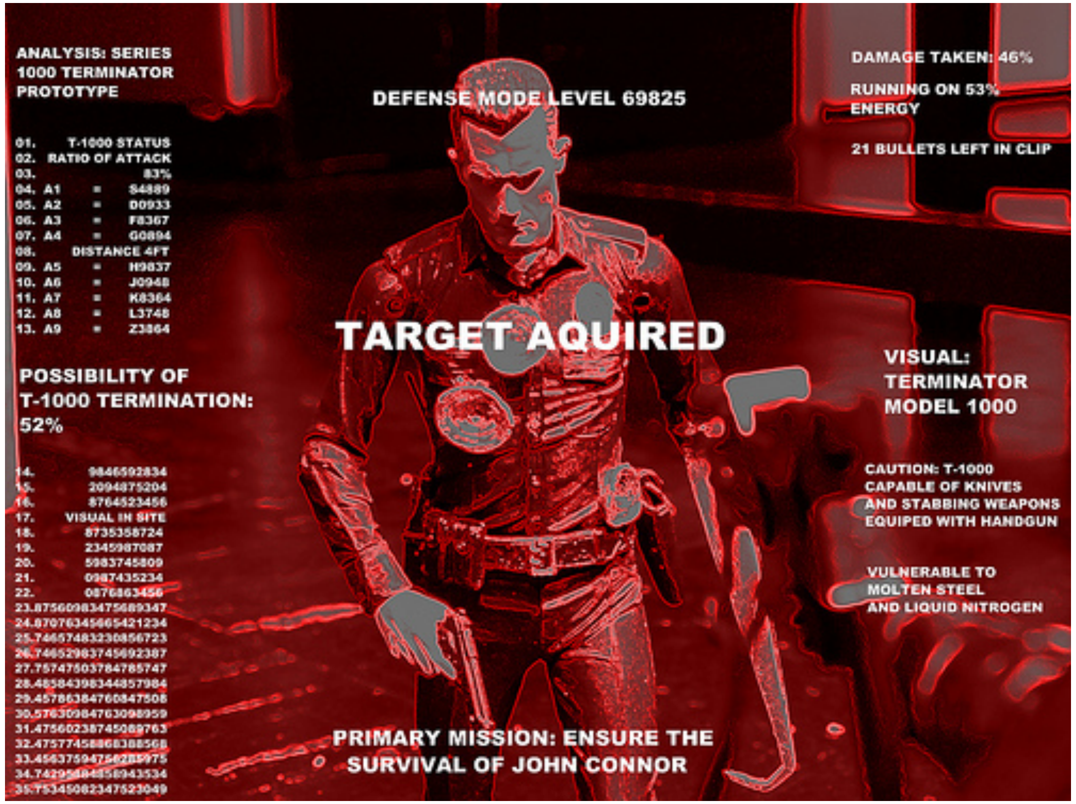
Reliable

Centre For MachineIntelligence

UNIVERSITY OF Southampton

AXA Research Fund

Who makes optimal decisions?



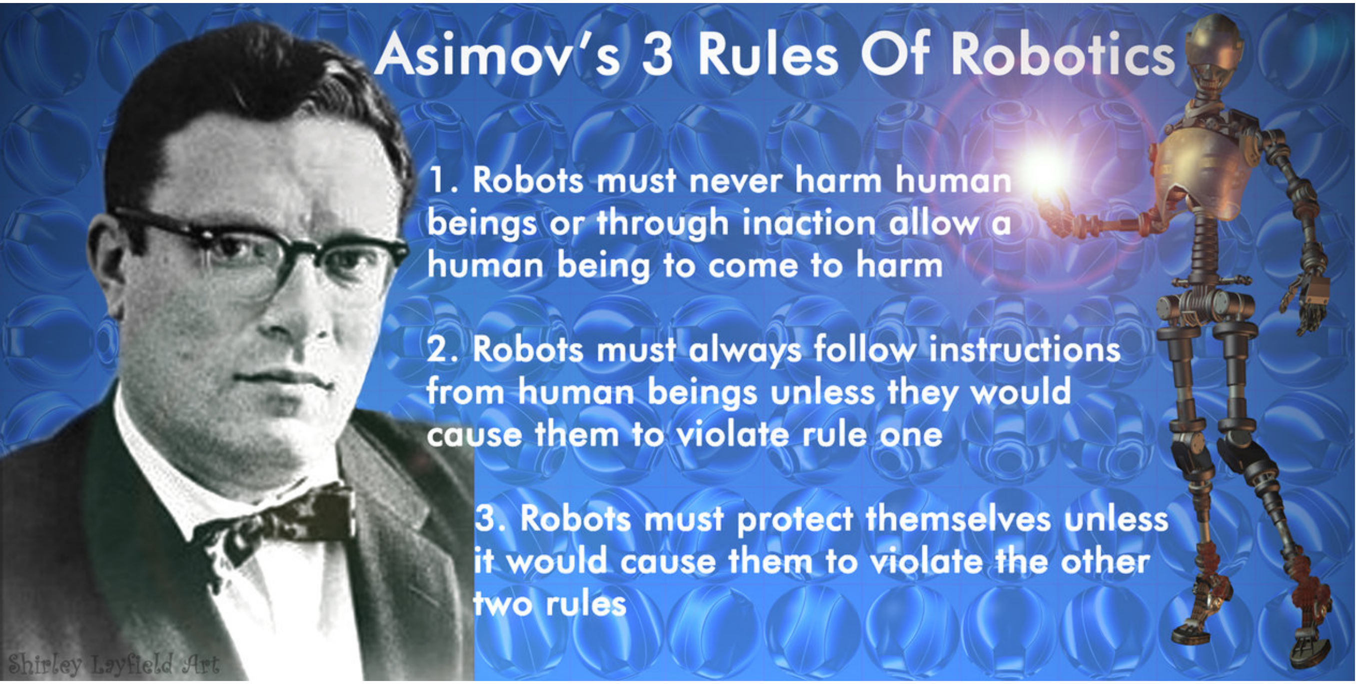
Centre For MachineIntelligence

UNIVERSITY OF Southampton

AXA Research Fund

Who is more precise?

Who makes the right decision?

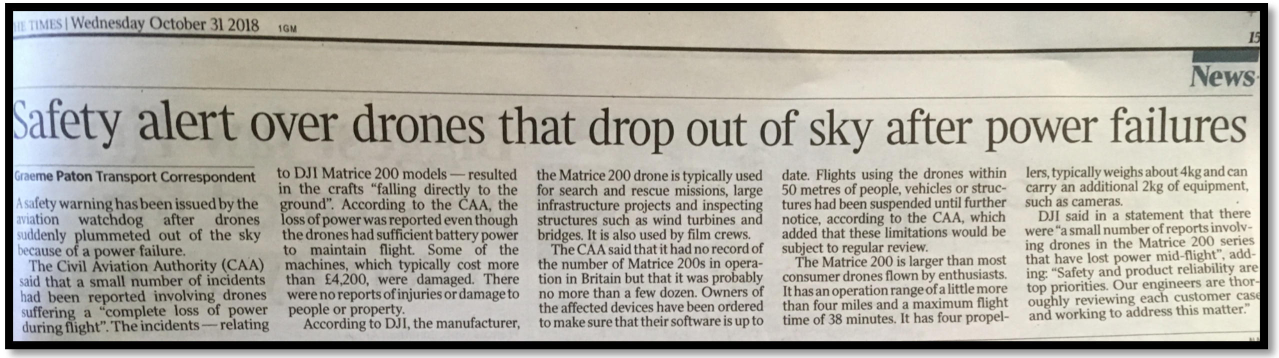


Centre For MachineIntelligence

UNIVERSITY OF Southampton

AXA Research Fund

Model Checking for Responsible Swarms


HE TIMES | Wednesday October 31 2018 1GM

15

News

Safety alert over drones that drop out of sky after power failures

Graeme Paton Transport Correspondent

A safety warning has been issued by the aviation watchdog after drones suddenly plummeted out of the sky because of a power failure.

The Civil Aviation Authority (CAA) said that a small number of incidents had been reported involving drones suffering a "complete loss of power during flight". The incidents — relating to DJI Matrice 200 models — resulted in the crafts "falling directly to the ground". According to the CAA, the loss of power was reported even though the drones had sufficient battery power to maintain flight. Some of the machines, which typically cost more than £4,200, were damaged. There were no reports of injuries or damage to people or property.

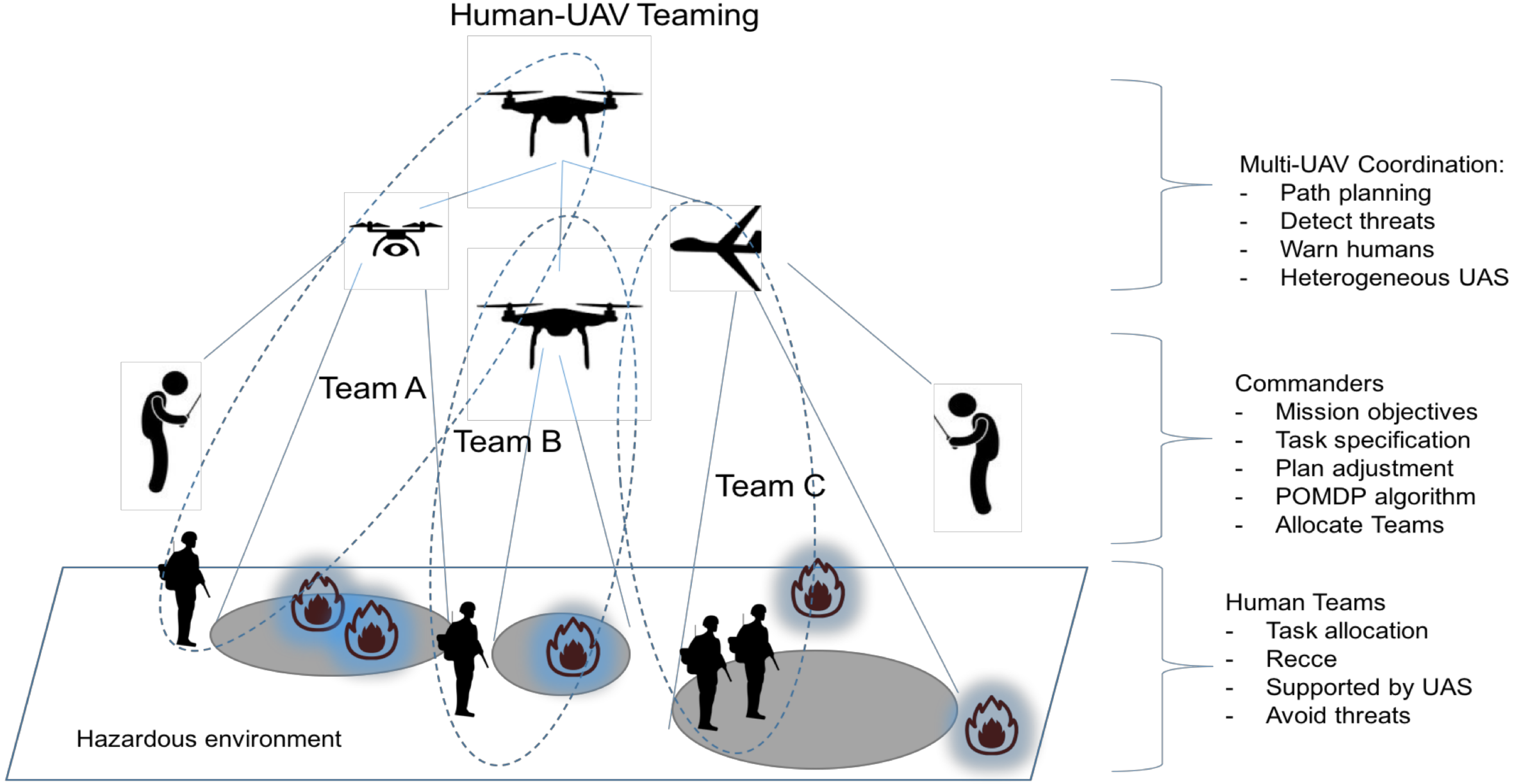
According to DJI, the manufacturer, the Matrice 200 drone is typically used for search and rescue missions, large infrastructure projects and inspecting structures such as wind turbines and bridges. It is also used by film crews.

The CAA said that it had no record of the number of Matrice 200s in operation in Britain but that it was probably no more than a few dozen. Owners of the affected devices have been ordered to make sure that their software is up to date. Flights using the drones within 50 metres of people, vehicles or structures had been suspended until further notice, according to the CAA, which added that these limitations would be subject to regular review.

The Matrice 200 is larger than most consumer drones flown by enthusiasts. It has an operation range of a little more than four miles and a maximum flight time of 38 minutes. It has four propellers, typically weighs about 4kg and can carry an additional 2kg of equipment, such as cameras.

DJI said in a statement that there were "a small number of reports involving drones in the Matrice 200 series that have lost power mid-flight", adding: "Safety and product reliability are top priorities. Our engineers are thoroughly reviewing each customer case and working to address this matter."

- Define Actions and Consequences
- Ethical behaviour and moral dilemmas
  - Control Dilemma:
    - Let the user fly or not?
  - Save life and Infrastructure:
    - Crash the drone or damage property
    - Do not harm humans
- Decide under uncertainty and dynamism



Centre For MachineIntelligence

Algorithms for Fair Load Shedding Problems

- Developing countries have an energy crisis
- Load shedding is essential
- Current load shedding techniques are not particular about fairness
- Can we use predictions of day-ahead consumption and supply to reduce unfairness?

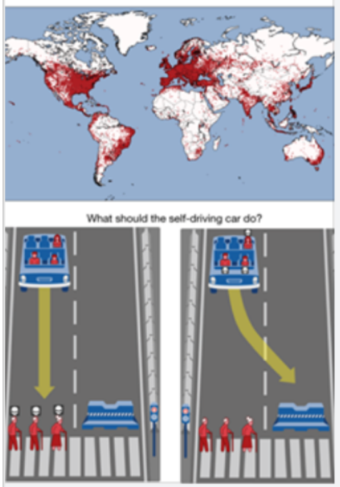
	Grouper Algorithm	Consumption Sorter Algorithm	Random Selector Algorithm	Cost Sorter Algorithm
Motivation	Minimize differences in disconnections	Minimize differences in disconnections & supplied electricity	Minimize differences in disconnections & supplied electricity	Minimize differences in comfort, disconnections & supplied electricity
Description	Random grouping of households & selection of group with least disconnections	Round-robin selection of households based on consumption	Round-robin selection of households in random order	Round-robin selection of households based on cost

AXA Research Fund

Open Questions

- There are no metrics for Responsible AI
- Methodological:
  - Algorithms (e.g., avoid training bias, privacy preserving)
  - Interfaces (e.g., avoid automation bias)
  - Organisations (e.g., guarantee safety and ethical outcomes)
- Challenges:
  - Modelling humans
  - Evaluating interfaces and interactions
  - Incentives to change
  - Dealing with Ethics

Awad et al., (2018) "The Moral Machine Experiment", Nature.



NATURE.COM

The Moral Machine experiment

Responses from more than two million people to an internet-based survey of attitudes towards moral dilemmas that might be faced by autonomous vehicles shed light on similarities