RGB ASSURANCE

Ironies of Automation with AI

Simon Connelly & Holger Becht
RGB Assurance

24 May 2019

Overview

- Motivation
- Literature Review
- AI - New technologies, new issues?
- Ironies of Automation with AI
- How can we design systems with AI?
- Conclusions

Motivation

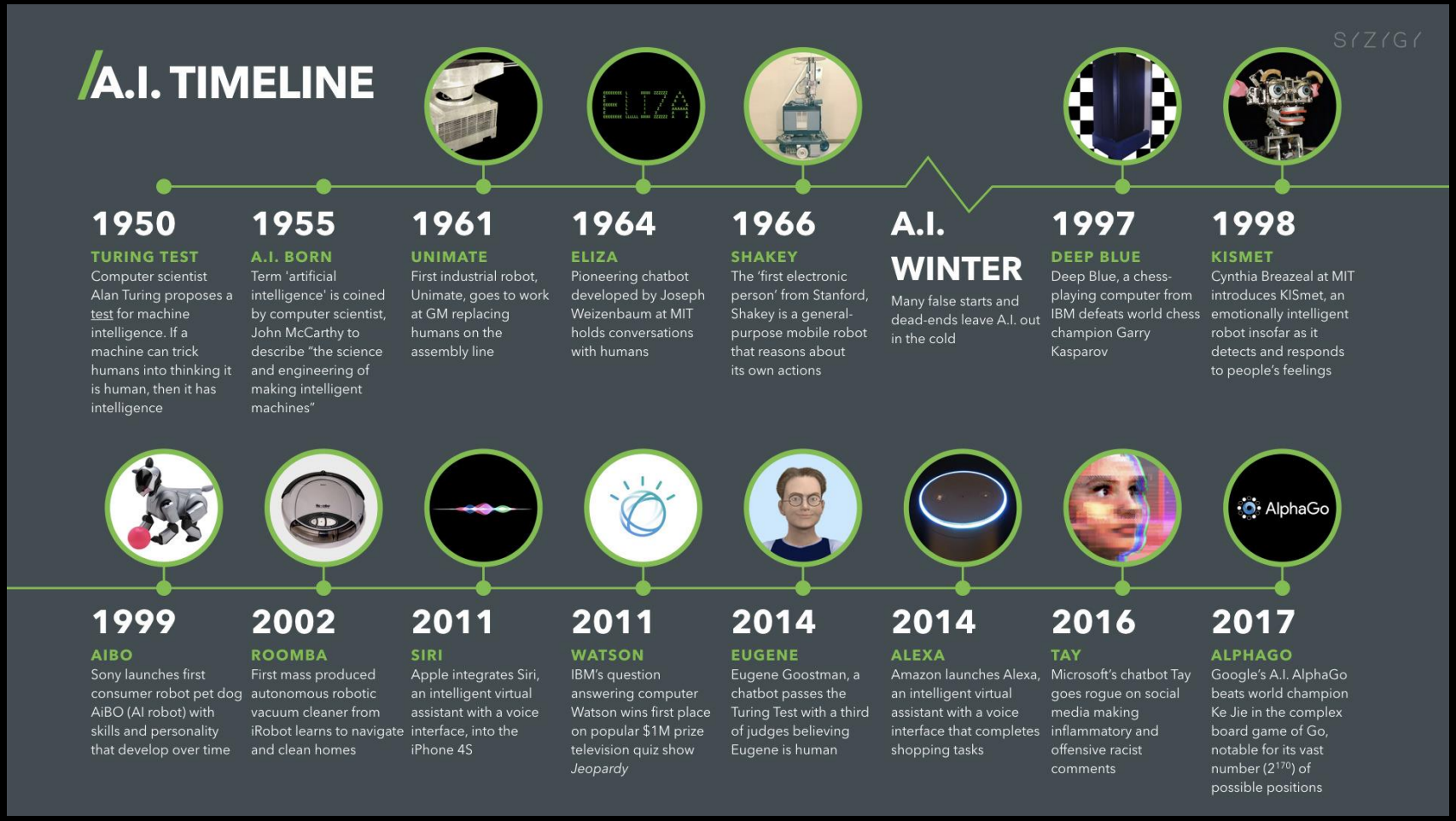
- Bainbridge argued that automation may expand rather than eliminate problems with human operators

  Lisanne Bainbridge (1983), *Ironies of Automation*
- Society is becoming more and more dependent on automation – still require human in-the-loop
- This talk
  - reflects on the original ironies of automation
  - discusses where and how the ironies persist with AI
  - discusses the implication on the role of the human operator
- *In this talk, we use the term "Automation" to include AI*

Ironies of Automation

- Classical Aim
  - to replace human manual control, planning and problem solving by automatic devices and computers
  - leaving the operator with responsibility for abnormal conditions and for on-line decision-making
- The irony
  - the more advanced or complex a control system is, the more crucial may be the contribution of the human operator

AI Timeline



Source: Dr Paul Marsden, digitalwellbeing.org

AI – New technologies, new issues?

- If AI is replacing human operators
  - what is the role of human operators?
  - is the operator capable and equipped to mediate these systems?
  - does the operator understand the capabilities and state of the system?

Boeing 737 MAX 8 – MCAS

- Two recent aircraft crashes
  - Lion Air Flight 610 (29 Oct 2018)
    - Crashed 12 minutes after takeoff
  - Ethiopian Airlines Flight 302 (10 Mar 2019)
    - Crashed 6 minutes after takeoff
- At the centre of the investigation is the MCAS
  - Manoeuvring Characteristics Augmentation System (MCAS), a **new** stall protection system

Boeing 737 MAX 8 – MCAS

- During take-off / on the ground
  - abnormalities in the altitude and the airspeed
  - problems with one of the angle of attack (AoA) sensors
- Shortly after takeoff
  - issues involving altitude and airspeed
  - commanded automatic nose-down trim via the MCAS
  - flight crew repeatedly commanded nose-up trim
- Investigation (ongoing)
  - believed a malfunction in the AoA sensors could lead the MCAS to believe that the aircraft is stalling, causing it to automatically initiate a dive
  - Limited information about the MCAS available to pilots in the manuals
  - Pilots were unaware of what was going on, they did not know how to override the MCAS, or even that it was the MCAS causing the problem

Concerns highlighted by 737-MAX 8

- Key issues can be seen through this example:
  - Operational uncertainty
  - Performance Predictability
  - Feedback to the user
  - User Alienation
- Key to addressing these is coordination

Dealing with Uncertainty

- Operators being unaware of what the automation system is doing, or how to mediate it, is not a new problem
  - AI appears to exacerbate the uncertainty and unpredictability
- This uncertainty adds additional complexities to human and AI integration
- Assurance of AI (e.g. adaptive) systems is challenging
  - run-time changes introduce a high degree of uncertainty
  - *We will not cover assurance of AI in this talk; this topic is covered by several other presentations*

Predictability

- Behaviour of adaptive systems may not be predictable to an operator
  - May hinder situation awareness by taking them out of the loop
  - May hinder development of correct mental model
  - Create another team member

Billings and Woods (1994); Woods (1996)

- Minimising automation surprises is critical to the success of implementing automation

Feedback

- Feedback is a critical component of well-designed automation
- Feedback is the only way the state of the system is communicated
- A lack of appropriate feedback is at the heart of many of the issues raised regarding the implementation of automation

User Alienation

- Out-of-the-loop (OOTL) performance creates alienation
- Operators are both slow to detect and correct problems

  Wickens & Kessel (1979); Young (1969); Endsley and Kiris (1995)
- Optimizing the automation's performance alone does not lead to gains in overall system performance

  McBride et al. (2015)
- Operators have better decision making accuracy when they have a better situational awareness and they remain competent in making those decisions

Coordination

- Operator and system must both understand what the other can and cannot do
- Coordination across agents in the system is at least as important as the performance of the individual agents taken in isolation
- Coordination design needs to consider behaviours during abnormal conditions

(Woods & Patterson, 2000)

- Ideally, the system and human should be considered as a joint cognitive system

Building Trust

- Building trust is important for effective coordination
- Beyond the scope of this talk
  - Transparency & feedback
  - Provenance
  - Understanding the capabilities

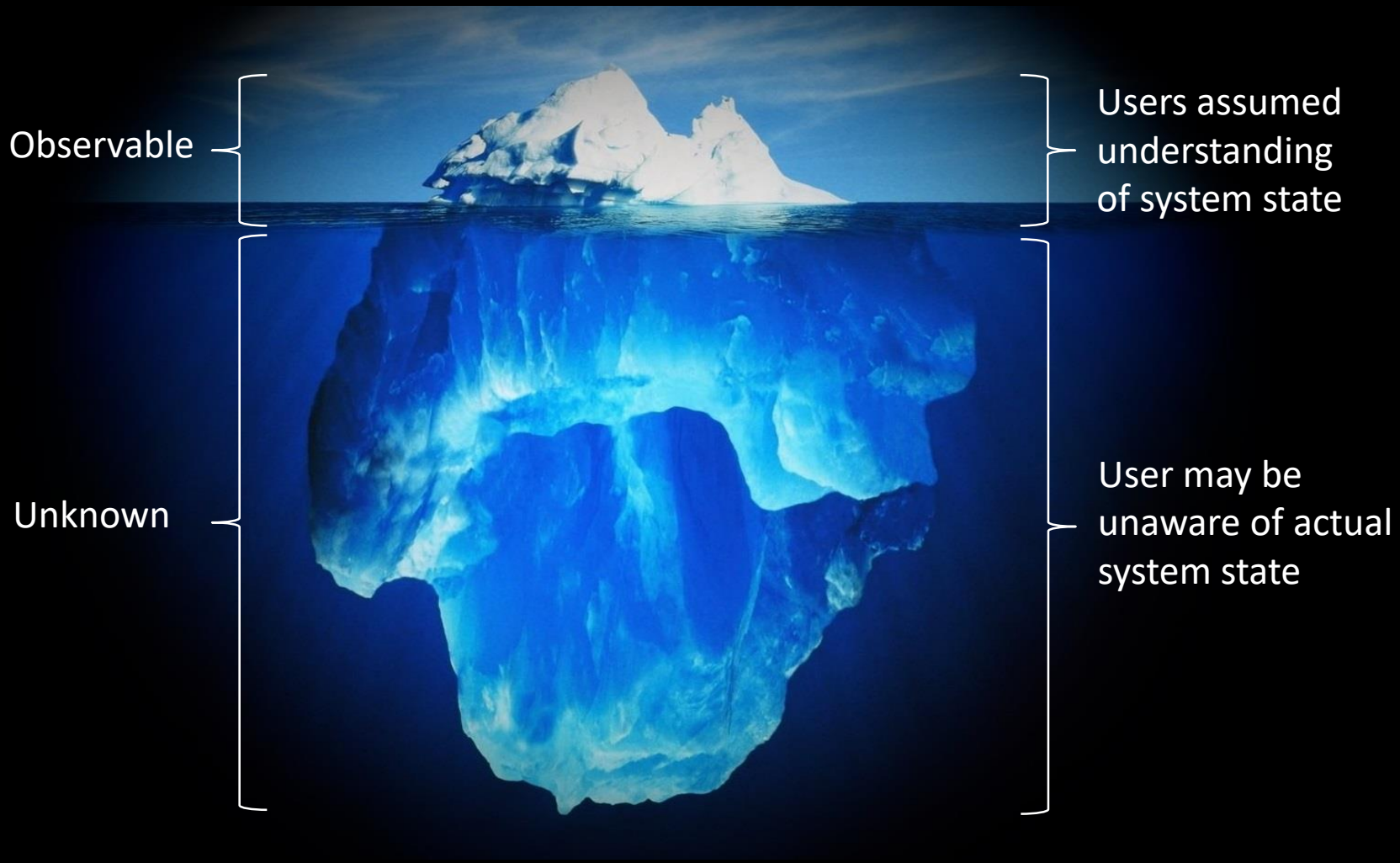
Irony of Automation with AI

- Still going strong:
  - The more we depend on technology and push it to its limits, the more we need highly-skilled, well-trained, well-practised people to make systems resilient, acting as the last line of defence against the failures that will inevitably occur
  - Expecting operators to supervise highly adaptive systems without adequate situational awareness and understanding of the automation is futile

How can we design Systems with AI?

- Artificial intelligence technology can result in artificial stupidity if it's poorly designed, implemented, or adapted
  - Need to ensure it's designed to help and coordinate with humans
  - Application of Human “Centred Engineering” and “Human-Centric AI”
    - Coordination is key – understanding who does what and what role the operator(s) play

Users understanding of the automation system state



Conclusions

- Critical for Human-Centric AI
  - Clearly define and understand the role of the human
    - Joint cognitive system view of teamwork and collaboration
  - Feedback and Observability
  - Understanding of system state and modes
  - Behaviour must be predictable and trustworthy

Questions