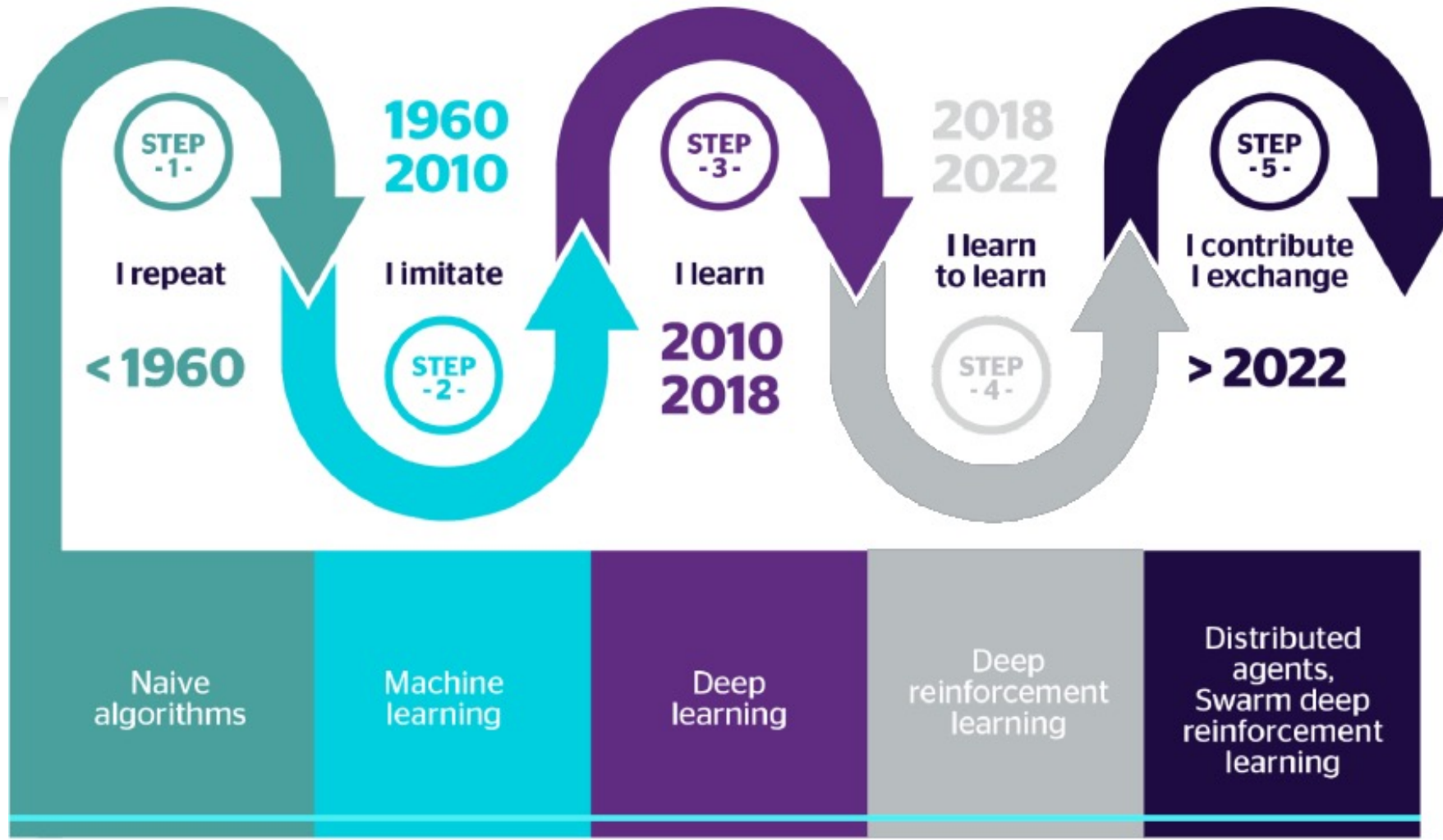# Responsible AI

2021 – 10 – 26
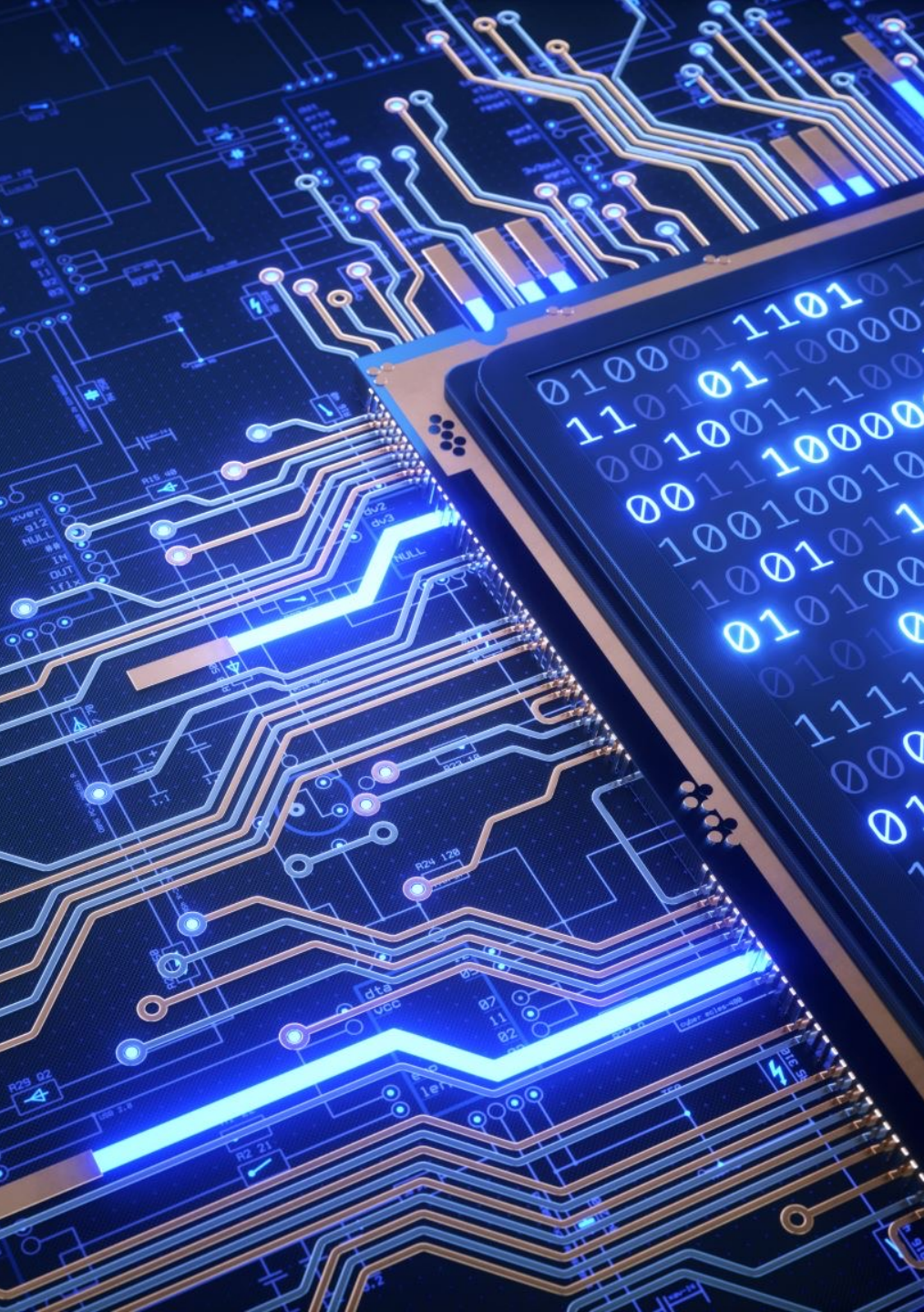
The start

- It all started with the pioneering work A. M. Turing (1950) Computing Machinery and Intelligence. *Mind 49:* 433-460.

- His logical framework was: "Humans use available information as well as reason in order to solve problems and make decisions, so why can't machines do the same thing?"

- During that period AI was discussed among scientists, mathematicians and philosophers.

- Today this discussion is also present in media, politics, business sector etc.

The timeline

| Naive algorithms | Machine learning | Deep learning | Deep reinforcement learning | Distributed agents, Swarm deep reinforcement learning |

The present

- AI is today is ubiquitous across diverse application fields and domains.

- There are high expectations from AI today in each of these fields and application domains

- To match these expectations there are nowadays 400+ policy documents that address the ethical issues of AI.

Some examples

AI and smart traffic lights could transform your commute

By Ana Moreno, CNN
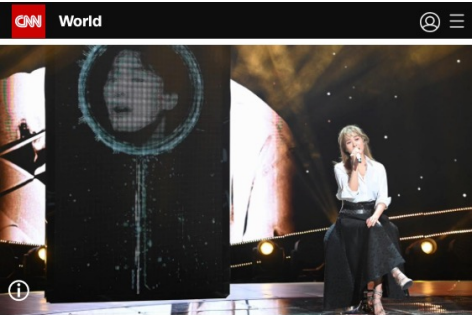Updated 1532 GMT (2332 HKT) December 29, 2020

Övervikt och är vad expert

Trump SPAC much as 1,65 was announc

World

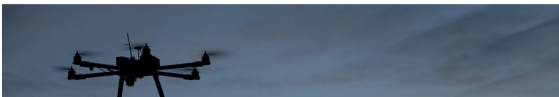South Korea has used AI to bring a dead superstar's voice back to the stage, but ethical concerns abound

By Gawon Bae, CNN
Updated 0234 GMT (1034 HKT) January 26, 2021

Seoul (CNN) — For the first time in 25 years, the distinctive vocals of South Korean superstar Kim Kwang-seok will be heard on national television singing new material.

WIRED
BACKCHANNEL   BUSINESS   CULTURE   GEAR   IDEAS   SCIENCE   SECURITY

WILL KNIGHT   BUSINESS   05.10.2021 08:00 AM
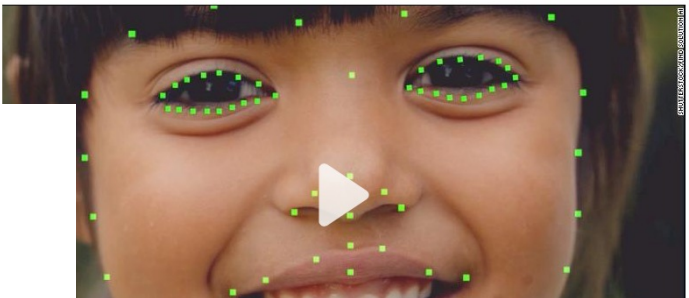
The Pentagon Inches Toward Letting AI Control Weapons

Drills involving swarms of drones raise questions about whether machines could outperform a human operator in complex scenarios.

BUSINESS EVOLVED

How AI that reads emotions is changing the online classroom

A Hong Kong company has developed facial expression-reading AI that monitors students' emotions as they study. They say the technology could make the virtual classroom as good as — or better than — the real thing.

Source: CNN

CODED BIAS
AVAILABLE TO STREAM APRIL 5TH ON
NETFLIX

"THE BEST OF SUNDANCE"

"'CODED BIAS' SERVES AS BOTH A WAKE-UP CALL AND A CALL TO ACTION"

"THE MOST CLEAREYED OF SEVERAL RECENT DOCUMENTARIES ABOUT THE PERILS OF BIG TECH"
The New York Times

codedbias.com/screen

Arts

Rembrandt's 'Night Watch' on display with missing figures restored by AI

Published 24th June 2021

Facebook is testing AI to get you to stop fighting in its groups

By Rachel Metz, CNN Business
Updated 2139 GMT (0539 HKT) June 16, 2021

TOP STORIES

How space researchers knew that 90-year-old William Shatner didn't...

Top oil exporter Saudi Arabia targets net zero emissions by 2060

Facial recognition tech has been widely used across the US government for years, a new report shows
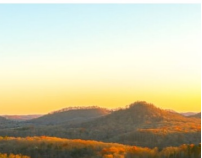
By Rachel Metz, CNN Business
Updated 1715 GMT (0115 HKT) June 30, 2021

World

Is this AI-powered indoor farm the future of agriculture?

Updated 0744 GMT (1544 HKT) October 6, 2021

NEWS

Facebook Sorted Countries Into 'Tiers' to Decide When to Interfere With Election Issues

BY ANDREW STANTON ON 10/25/21 AT 11:10 AM EDT

Facebook Whistleblower Claims 'Anger And Hate Is Easiest Way To Grow' On Platform

Topics

AI-Savvy Criminals Clone Executive's Voice in $35 Million Deepfake Bank Heist

By Vanessa Bates Ramirez - Oct 20, 2021   4,658

Power and Resposibility

Some perspectives

Ethics and design perspective

Algorithmic perspective

Control perspective

User experience perspective

Principled AI


Ethical concerns are catalogued in the Berkman Klein Center report that offers ethical principles in eight categories*: privacy*, *accountability*, *safety and security*, *transparency and explainability*, *fairness and non-discrimination*, *human control of technology*, *professional responsibility*, and *promotion of human values*.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, (2020-1).

Value Sensitive Design

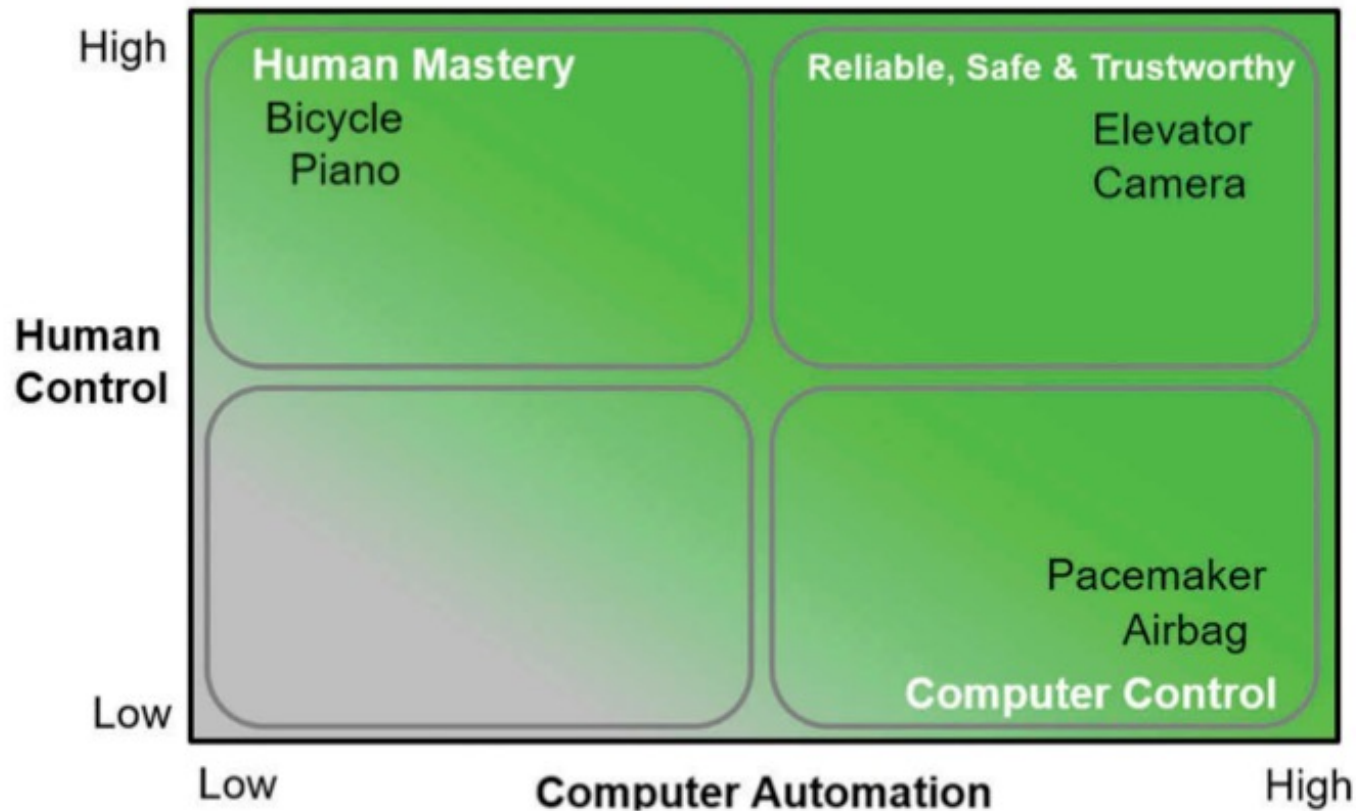Value sensitive design (VSD) is an established method for integrating values into technical design.
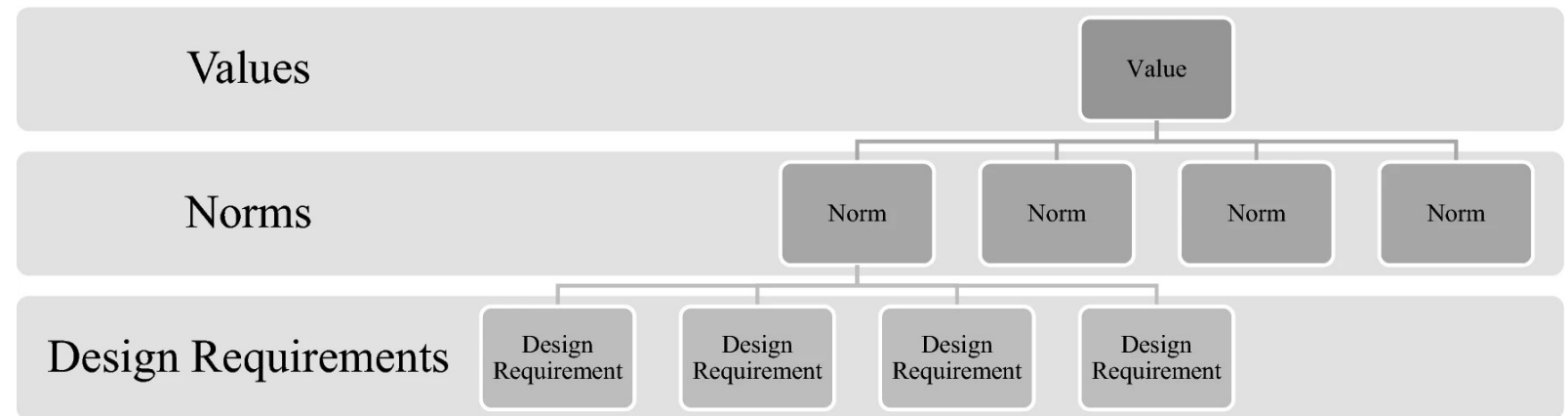
Control

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495-504.

User experience

- The lack of understandability and explainability is defined by researchers as a crucial feature that affects the practical deployment of ML and AI solutions.

- The need for understanding and ability to explain AI (Adadi and Berrada, 2018), stems from four main reasons:
  - Explain to justify
  - Explain to control
  - Explain to improve
  - Explain to discover

Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6, 52138–52160 (2018)

Our Responsibilities?

- Our data create the base for algorithms

- Algorithms are the base of AI systems

- AI system exercise decision

- Decisions have social implications

- Who is the responsible for these social implications?

Discussions

- What is our responsibility as individual users for responsible AI?

- What is our responsibility as designers for responsible AI?

- What is our responsibility as researchers for responsible AI?

- What is our responsibility as a society for responsible AI?

Workshop

- Division into 4 groups

- Each group gets one of the discussion question and elaborates (ca 15 min.)

- We use Jamboard to document the discussions
https://jamboard.google.com/d/1R0r-SvfgVqco7Ak3uCiFA-YCTNqZSrfho6YeiYOMey8/edit?usp=sharing

- We summarize together our discussions