



AI could kill all humans in next decade, warn experts: but how seriously should we take them?

Alarming warnings from industry insiders increase pressure for curbs on artificial superintelligence



9/9/2026



A quick question.

In the last month, has a software helped decide something important in your life, your insurance, a medical result, a university course?

Keep your hand up only if you could get the exact reason it decided that about you, not a customer-service answer. The real reasoning.

Society 5.0: Challenges for Industrial Management

A critical look at the world's boldest plan for technology and society

ACT I • THE PROMISE — WHERE THE NAME COMES FROM

Five societies, one story



Japan calls this the fifth stage of human civilization, the same size of claim as the invention of farming.

What Japan actually promised

“A human-centred society that balances economic growth with solving social problems through a system that closely joins the digital world and the physical world.”



1 • Human-centred

Technology should serve people, not the other way around.



2 • Problem-solving

Aimed at real problems: ageing, energy, disaster, healthcare.



3 • Cyber-physical

Sensors, data, machines, and people joined into one non-stop system.

Three promises. Today, we test all three

Source: Cabinet Office of Japan, 5th Science and Technology Basic Plan (2016).

A promise born from necessity

Society 5.0 is not a luxury project. It answers a demographic emergency with no historical precedent.

26.9%

of people in Japan are aged
65 or older, highest in the
world

-12%

fall in the working-age
population since its 1995
peak

11M

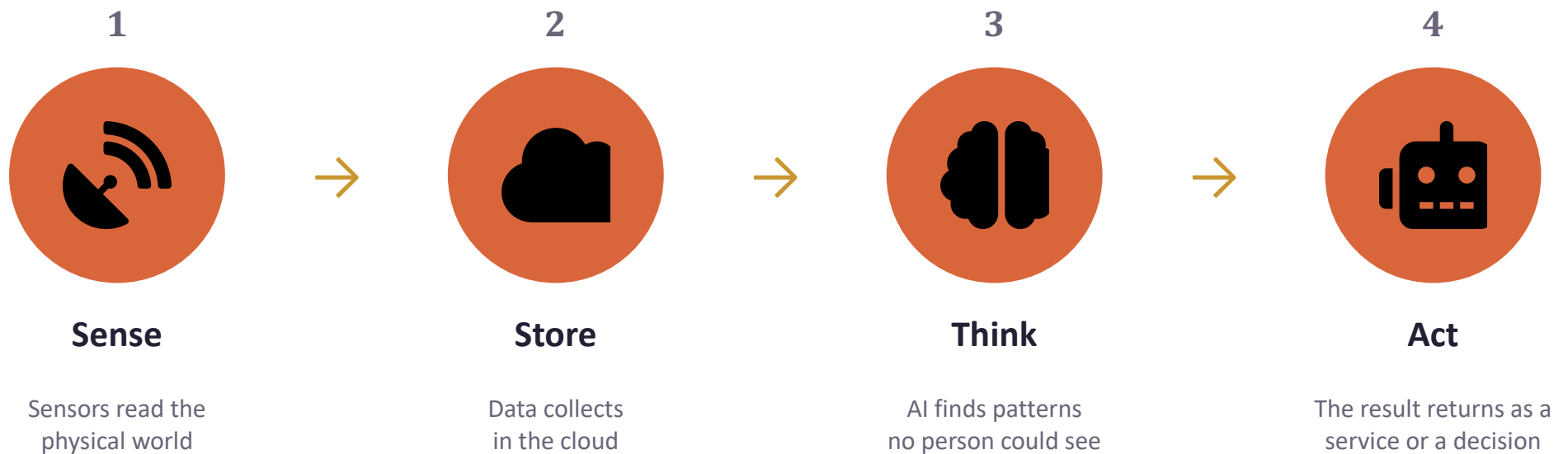
projected worker shortfall
by 2040

1.4

births per woman , far
below the replacement
level

2016 Data

The engine behind the promise



The same loop repeats, everywhere, without stopping.

ACT I • THE PROMISE

The promise is genuinely beautiful



Older people live independently, cared for by robots. Never alone.



Illness is found before they appear.



Cities act before the earthquake, the flood, the accident.



Hard, repetitive work goes to machines, so people can create.

Japan even ties this promise directly to the United Nations' Sustainable Development Goals.

Even Japan changed its mind

2016

“...a system that closely joins the digital world and the physical world.”

A claim about architecture.

2021

“...a society that is sustainable and resilient against threats and uncertainty.”

A claim about survival.

¥60T

committed to public research — a record for Japan

13th→3rd

target world rank for research within 10 years

2026

the year the 7th Basic Plan took effect — no longer centred on Society 5.0

Somewhere between 2016 and 2021, Japan stopped describing a destination and and began describing itself as resilient in the face of threats

Source: Cabinet Office of Japan, 6th (2021) & 7th (2026) S&T Basic Plans; independent reporting confirms the 2026 shift away from Society 5.0.

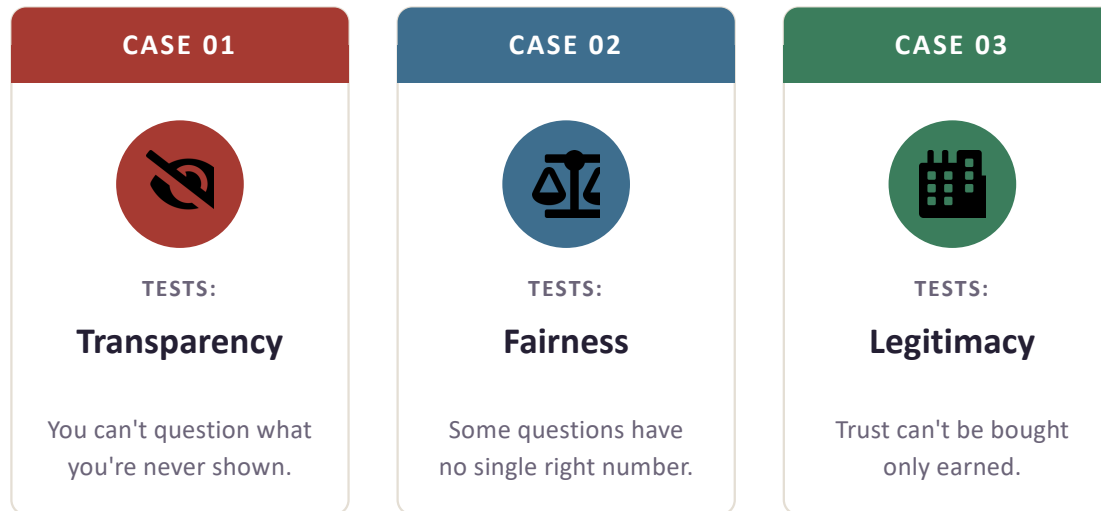
The gap

In every case that follows, the technology worked.
The institution around it did not.

ACT II • THE TEST — WHAT TO LISTEN FOR

Three cases, three different ways the promise broke

What stays constant: sense data, store, decide, act and the same promise, that the loop can be trusted with a real decision.





CASE 01 • THE TRANSPARENCY GAP

SyRI, and the scandal behind it

“Not enough transparency and proportionality.” — District Court of The Hague, February 2020

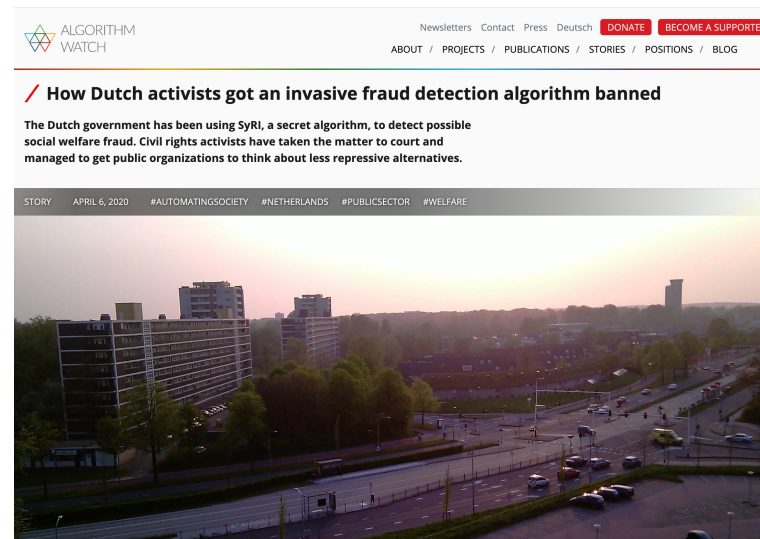
The Dutch government used an algorithm, SyRI, to flag people suspected of benefit fraud. A related system wrongly accused tens of thousands of families, many from immigrant backgrounds, of child-benefit fraud, demanding repayment that ruined lives.

The promise: "human-centred" means citizens can see and question what a system decides about them.

The lesson: when they cannot, injustice can move faster than any appeal can reach.

Source: <https://algorithmwatch.org/en/syri-netherlands-algorithm/>

ACT II • THE TEST





CASE 02 • THE FAIRNESS GAP

COMPAS. Not a bug, a proof.

The promise: an algorithm can compute a fairer answer than we can.

The defence — Northpointe (the maker): the tool was calibrated. A given score meant the same real risk for every group. True.

The counter — ProPublica (2016): the tool was twice as likely to wrongly flag Black defendants as high-risk. Also true.

The proof — Kleinberg, Mullainathan & Raghavan (2017): when two groups differ in their real rates, a test cannot be both calibrated and equal in its errors. Not a bug a theorem.

The lesson: fairness is not something engineering can compute.
It is a choice between incompatible ideas of what “fair” means.

Source: Angwin et al., ProPublica (2016); Kleinberg, Mullainathan & Raghavan, ITCS (2017).

ACT II • THE TEST

Inherent Trade-Offs in the Fair Determination of Risk Scores

Jon Kleinberg^{*} Sendhil Mullainathan[†] Manish Raghavan[‡]

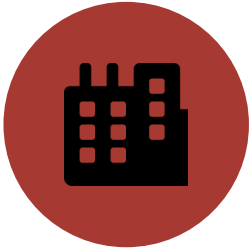
Abstract

Recent discussion in the public sphere about algorithmic classification has involved tension between competing notions of what it means for a probabilistic classification to be fair to different groups. We formalize three fairness conditions that lie at the heart of these debates, and we prove that except in highly constrained special cases, there is no method that can satisfy these three conditions simultaneously. Moreover, even satisfying all three conditions approximately requires that the data lie in an approximate version of one of the constrained special cases identified by our theorem. These results suggest some of the ways in which key notions of fairness are incompatible with each other, and hence provide a framework for thinking about the trade-offs between them.

1 Introduction

There are many settings in which a sequence of people comes before a decision-maker, who must make a judgment about each based on some observable set of features. Across a range of applications, these judgments are being carried out by an increasingly wide spectrum of approaches ranging from human expertise to algorithmic and statistical frameworks, as well as various combinations of these approaches.

15807v2 [cs.LG] 17 Nov 2016



CASE 03 • THE LEGITIMACY GAP

Sidewalk Labs, Toronto

“Almost unlimited money. Excellent engineers. It still could not win public trust and it could not buy that trust either.”

Google's parent company, Alphabet, tried to build a smart district from scratch on Toronto's waterfront. The project ended in 2020 after years of public opposition over who would own the data it collected. Pandemic-related uncertainty played a real role too.

The promise: "human-centred" means built for citizens.

The lesson: trust is not something engineering can manufacture; it has to be earned.

Why Did Sidewalk Labs's Quayside Project Fail? The Practical and Institutional Difficulties of Smart City Practice

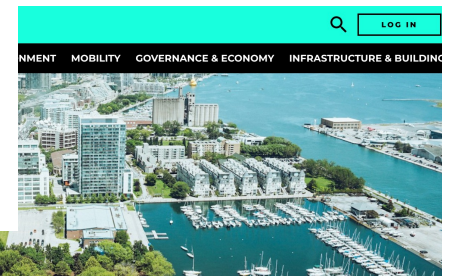
May 2021

Conference: The Environmental Design Research Association (EDRA) 52 Detroit • At: Detroit, Virtual

Authors:



Jun Wang
Georgia Institute of Technology



This article is also available [here](#) in Spanish.

SIDEWALK TORONTO, THE VISION BEHIND GOOGLE'S FAILED CITY

EDRA 52 DETROIT 2021

Source: <https://www.tomorrow.city/sidewalk-toronto-the-vision-behind-googles-failed-city/>

Why did Sidewalk Lab's Quayside project fail? The practical and institutional difficulties of smart city practice. Jun W. Georgia Institute of Technology. Conference: The Environmental Design Research Association (EDRA) 52 Detroit (2021)

ACT II • THE TEST

We built the digital-physical layer on credit and promised the human institutions would arrive later.

They never did.

In every one of these three cases: a technical system that worked, and no institution ready to answer for what it did.

Whoever writes the rules, writes the future



JAPAN

Society 5.0

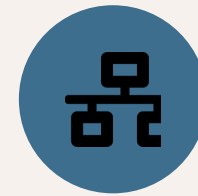
Government + enterprises. Solve social problems through the sense-store-think-act loop.



EUROPE

Industry 5.0

Regulator-led. Worker wellbeing and sustainability ahead of shareholder profit alone.



PLATFORM

“Default 5.0”

“Nobody” designed this one. Whoever owns the data and the models sets the terms.

Sources: Japan Cabinet Office (2016); European Commission, Industry 5.0 (Breque, De Nul & Petridis, 2021).

Two cities, one technology



CITY A

Sensors everywhere. Data owned by citizens. Every algorithm can be checked and switched off.

Freedom, extended.



CITY B

The same sensors everywhere. Data owned by whoever runs the platform. No one outside can see how decisions are made.

Control, perfected.

Identical hardware. Opposite societies. The difference is not the technology it is who governs it.

The problem was never putting humans at the centre.

The problem is which human, a statistic, or a real person.

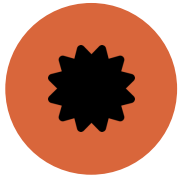
In every case you saw, that statistic collided with a real person who had no voice to object.

Not spectators the people who sign



1 • Teach it properly

Treat the question of who answers for a model with the same rigour as engineering fundamentals, not an optional class.



2 • Write it into accreditation

Make institutional and ethical competence a requirement for a degree to be approved not a nice suggestion.



3 • Reward the follow-up study

Fund and reward the unglamorous study that takes years, not months. What the system actually did in practice, not only the exciting first pilot.

The future is not forecast

It is authored

We shouldn't ask what technology will do to us

We should ask what we will demand of it.



Who holds the pen?