

AI, Robotics, and the Algorithmification of Ethics in South Korea

Chihyung Jeon (Sci & Tech Policy, KAIST)
cjeon@kaist.edu

Digital Asia Hub Workshop on AI and Ethics
16 December 2016

News articles on “artificial intelligence” in Korea



인공지능



관련 [인공지능의 편견](#) [4차 산업혁명](#) [인공지능 사라질 직업](#) [인공지능 이란](#) [인공지능이 우리 사회에 미치는 영향](#) [인공지능의 장단점](#)
[인공지능 관련주](#) [인공지능 부정적 영향](#) [인공지능 로봇](#) [인공지능의 위험](#) [인공지능 프로그래밍](#) [인공지능 관련학과](#)

통합검색

이미지

동영상

뉴스

블로그

카페

트위터

웹문서

더보기 ▾

선택된 옵션 2012.02.01~2016.12.31 ×

2012.02.01

2016.12.31

검색



AlphaGo Shock

News articles on “artificial intelligence + ethics” in Korea



인공지능 윤리



- 통합검색
- 이미지
- 동영상
- 뉴스**
- 블로그
- 카페
- 트위터
- 웹문서
- 더보기 ▾

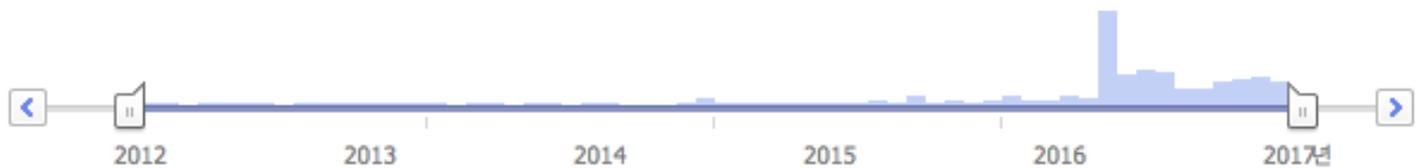
관련 [인공지능 윤리 문제](#) [인공지능 판단](#) [인공지능 윤리의식](#)

선택된 옵션 2012.01.01~2016.12.31 ×

2012.01.01

2016.12.31

검색



AlphaGo Shock

How and why are the critical moments of decision in AI-based systems *framed* as a problem of ethics? (e.g. the “trolley problem” in self-driving cars)

Why is the *full autonomy* of AI systems considered inevitable and desirable, with the consequence of equating the decisions made by an AI system with those made by a human operator?

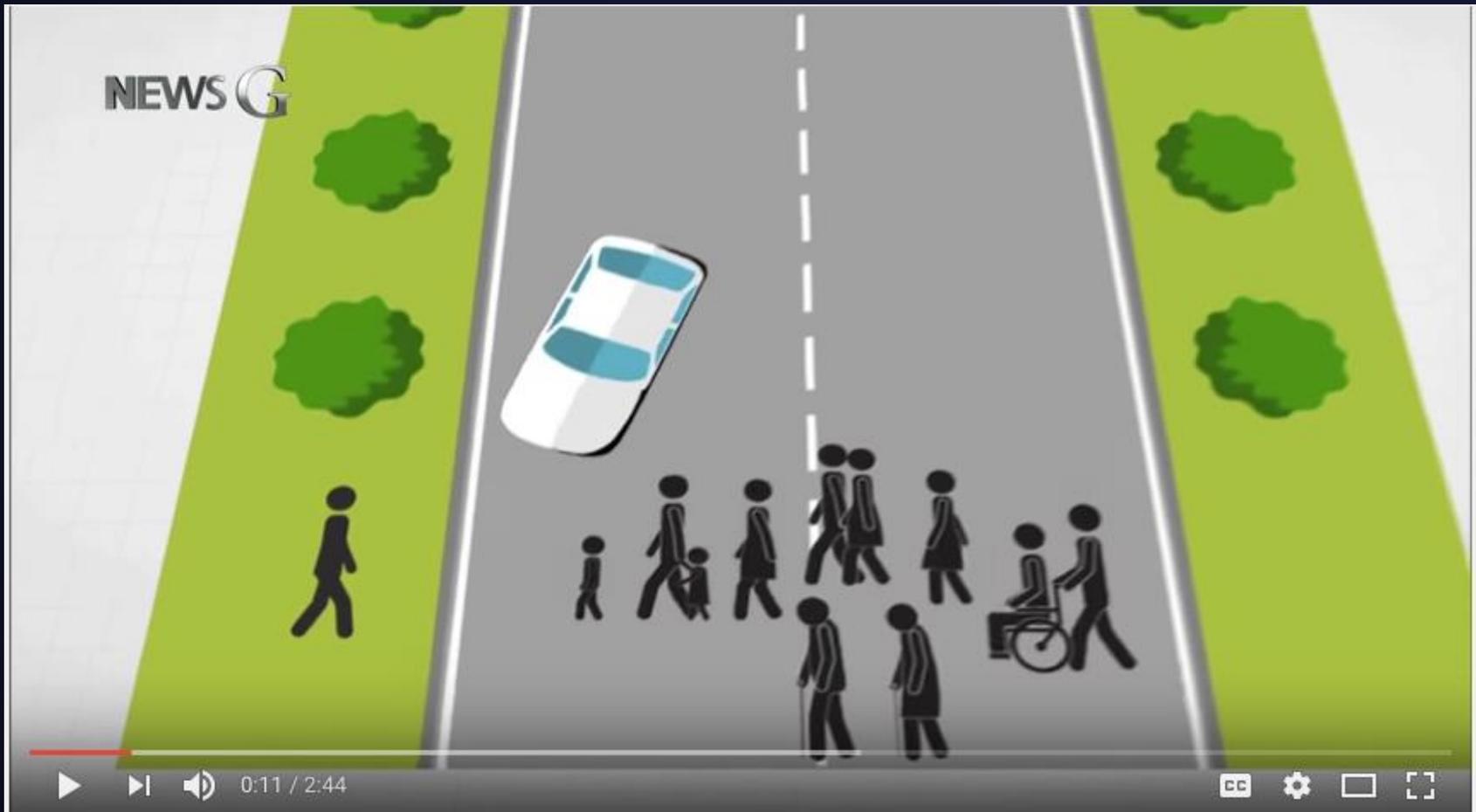
Where does ethics exist in/around AI systems?

How to study AI and ethics from STS perspectives?

Popularity of the trolley problem (in Korea)



Popularity of the trolley problem (in Korea)



[뉴스G] 인공지능이 '도덕'을 배울 수 있을까?

EBS News



EBS 뉴스

Subscribe 4,630

2,420 views

“Can (anthropomorphized) AI learn morality?”

NEWS G



▶ ⏪ 🔊 0:51 / 2:44

CC ⚙️ 📺 🗉

[뉴스G] 인공지능이 '도덕'을 배울 수 있을까?

EBS News



EBS 뉴스

▶ Subscribe 4,630

2,420 views

Assumptions found in AI ethics scenarios (in Korea)

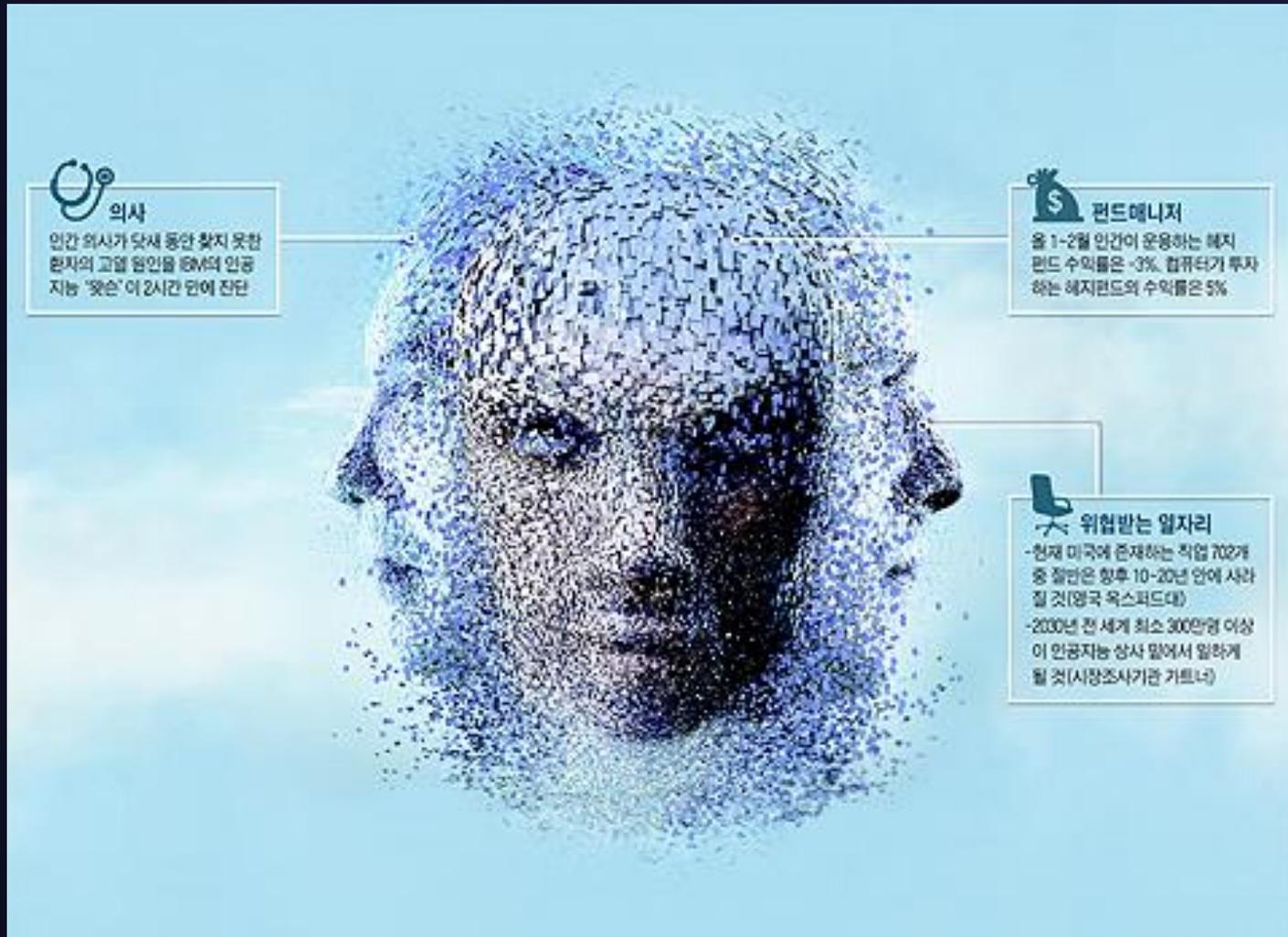
1. Most (hypothetical) situations considered in AI ethics discussion are offered as *instantaneous* events, in which critical life-and-death decisions have to be made in less than a second.

Assumptions found in AI ethics scenarios (in Korea)

1. Most (hypothetical) situations considered in AI ethics discussion are offered as *instantaneous* events, in which critical life-and-death decisions have to be made in less than a second.

2. AI that makes ethical decisions is frequently *anthropomorphized*, suggesting a full, one-on-one replacement of human operators by AI.

anthropomorphized AI on Korean media



“The AI Says...” *Chosun Ilbo*, 12 Mar 2016

anthropomorphized AI on Korean media



Joongang Ilbo, 8 Dec 2016

Assumptions found in AI ethics scenarios (in Korea)

1. Most (hypothetical) situations considered in AI ethics discussion are offered as *instantaneous* events, in which critical life-and-death decisions have to be made in less than a second.
2. AI that makes ethical decisions is frequently *anthropomorphized*, suggesting a full, one-on-one replacement of human operators by AI.
3. Ethics is presumed to exist *outside* and *prior to* the AI system, which may or may not be able to “learn” it.

Ethics discourse in AI narrows the operational time–frame of human–machine systems; focuses on AI systems, ideally, with no human intervention; and ignores a broader set of actors who are involved in a longer time–frame of development, regulation, and use. Ethical decision is assumed to be made instantaneously by the AI alone (with no human involvement). This tendency conceals the longer–term, probably more critical, decisions made by engineers, regulators, and users (made over a much longer period of time).

And there are few doubts that these “ethical problems” will be “solved” by prescribing or designing proper ethical rules into the system. And when it is solved, the assumption goes, it will be solved by ethical algorithms coded by ethical engineers, which turns ethics problems into another, albeit very difficult, coding problem. The process of coding may be discussed with humanists, social scientists, and regulators, but it remains as a coding problem. To the extent that “non-technical” issues of AI systems are framed as ethical problems, they are seen as curiously apolitical problems.

An STS (science, technology & society) Perspective

Life-and-death decisions on the road are not always made by a fully autonomous driver in less than one second.

The driver's decision/action is constrained by a range of mechanisms, artifacts, techniques, codes, laws, values, environment, etc. Ethics or morality is distributed across these material and non-material entities.

An STS (science, technology & society) Perspective

Ethics does not shift simply from a human to an AI.

Rather, ethics consists in the way that the human and the machine are configured vis-à-vis each other within the system. This is the case in any human-machine systems with or without AI.

“The myth of full autonomy”

“But the machine that operates entirely independent of human direction is a useless machine. Only a rock is truly autonomous (and even a rock was formed and placed by its environment).”

—David Mindell, *Our Robots, Ourselves* (2015)

We need to study:

not just “How ethics can be inserted into AI,”

but rather “How humans and AI machines are configured in a way that reflects and shapes particular ideas of ethics.”

More generally,
we need to study how AI works—not just within the
computer and inside the network.

But how AI works *in the world, with people, within
institutional, political, and cultural environments.*

i.e. sociological and anthropological study of
“AI in action.”

COMMENT

TREES A celebration of the poet of forest ecology, Oliver Rackham **p.314**



ENGINEERING On the many uses of rotation, from biology to business **p.315**

PALAEONTOLOGY Why are women in the field donning beards for equality? **p.316**

OBITUARY Deborah Jin, pioneer of ultracold physics, remembered **p.318**



Chicago police use algorithmic systems to predict which people are most likely to be involved in a shooting, but they have proved largely ineffective.

There is a blind spot in AI research

Fears about the future impacts of artificial intelligence are distracting researchers from the real risks of deployed systems, argue **Kate Crawford** and **Ryan Calo**.

On 12 October, the White House published its report on the future of artificial intelligence (AI) — a product of four workshops held between May and July 2016 in Seattle, Pittsburgh, Washington DC and New York City (see go.nature.com/2dx8rv6).

During these events (which we helped to organize), many of the world's leading thinkers from diverse fields discussed how AI will change the way we live. Dozens of presentations revealed the promise of using progress in machine learning and other AI techniques

to perform a range of complex tasks in everyday life. These ranged from the identification of skin alterations that are indicative of early-stage cancer to the reduction of energy costs for data centres.

The workshops also highlighted a major blind spot in thinking about AI. Autonomous systems are already deployed in our most crucial social institutions, from hospitals to courtrooms. Yet there are no agreed methods to assess the sustained effects of such applications on human populations.

Recent years have brought extraordinary

advances in the technical domains of AI. Alongside such efforts, designers and researchers from a range of disciplines need to conduct what we call social-systems analyses of AI. They need to assess the impact of technologies on their social, cultural and political settings.

A social-systems approach could investigate, for instance, how the app AiCure — which tracks patients' adherence to taking prescribed medication and transmits records to physicians — is changing the doctor-patient relationship. Such an approach ▶

“social-systems analysis”

Nature, 20 Oct 2016

Copyrighted Material

IN THE AGE OF THE SMART MACHINE

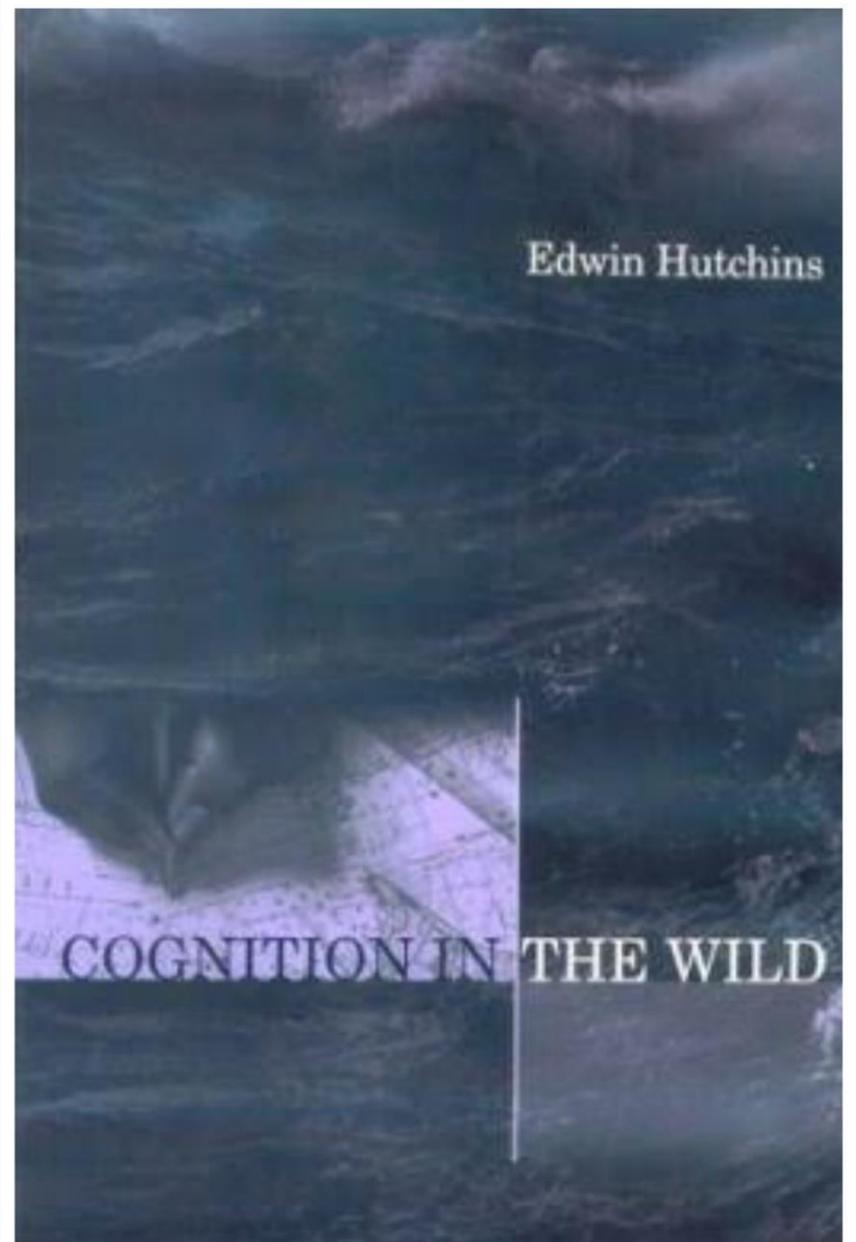
THE FUTURE OF
WORK AND POWER



SHOSHANA ZUBOFF

Copyrighted Material

1988



Edwin Hutchins

COGNITION IN THE WILD

1995

HUMAN-MACHINE RECONFIGURATIONS

PLANS AND SITUATED ACTIONS 2ND EDITION



LUCY A. SUCHMAN

CAMBRIDGE

2007



Our Robots, Ourselves

Robotics and the Myths of Autonomy

DAVID A. MINDELL

2015