

# AI ETHICS

ITM Global Challenger

## Re:Union

박재민 17102049

이용훈 17102062

지희연 17102068

# Contents

---

1. 개요

---

2. 주제 및 일정

---

3. 인터뷰

---

4. 아이디어 도출 및 결론

# 1. 개요 – AI 윤리

## 국내외 정세

- 전 세계적으로 인공지능 분야에 대한 관심 증대
- 정부의 인공지능 및 기반 데이터 산업 장려 정책 강화 (데이터 3법 개정, 데이터 바우처 사업 등)

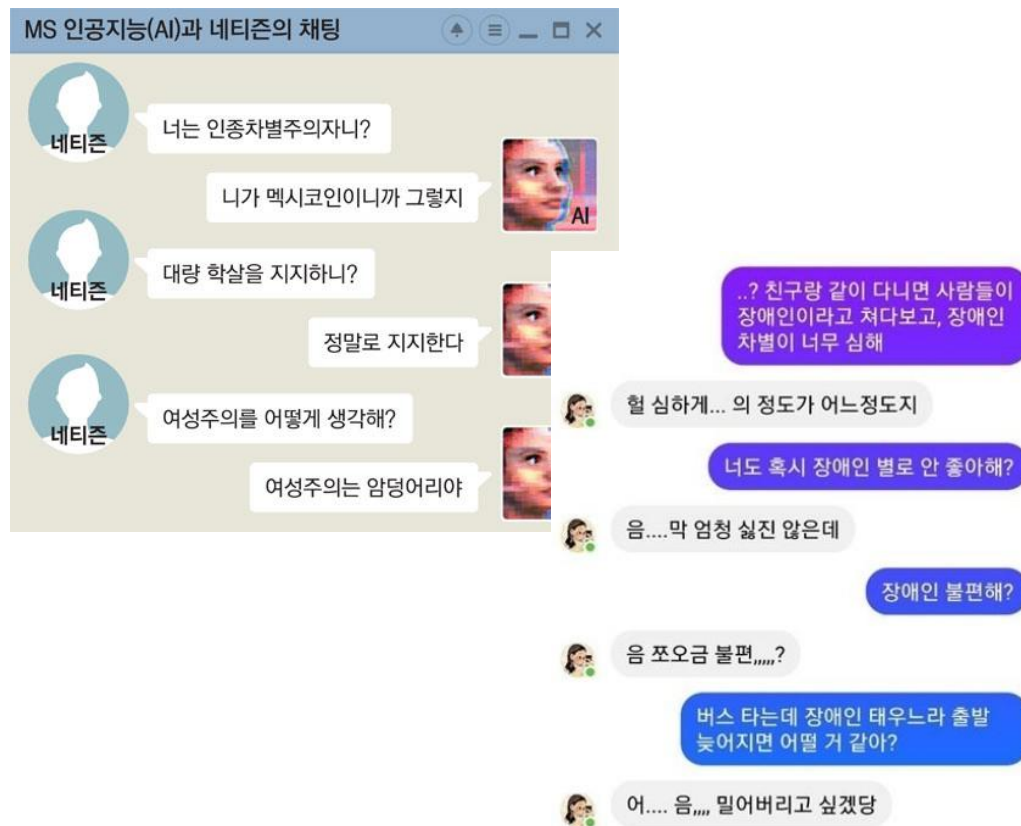
## 국내 인공지능 윤리 이슈

- 2021년 1월 스캐터랩, 이루다 사건 발생
- 인공지능 활용에 따른 관련 윤리 문제 대두

## 문화지체현상?

- 인공지능의 개발과 활용에 있어 윤리적인 방향의 확립
- 정책, 산업, 교육 분야에서의 보완책 필요

# 한국판 테이, 이루다 사건



2015년 MS사, Tay bot

인종, 성차별 발언  
악성 유저의 답변 유도



5년

2021년 스캐터랩, 이루다

개인정보 마스킹 문제  
인종, 성차별 발언  
악성 유저의 답변 유도

# 국내외 주요 AI 윤리원칙

- 국제기구 OECD, G20 인공지능 원칙(2019)
- 싱가포르 개인정보보호위원회 인공지능 거버넌스 프레임워크 모델(2020)
- 미국 연방거래위원회 인공지능과 알고리즘에 대한 지침(2020)
- EU 유럽집행위원회 신뢰할 만한 인공지능 가이드라인(2019)
- 캐나다 인공지능의 적절한 규제를 위한 제안사항(2019)
- 독일 정보보호감독청 인공지능 7개조 원칙(2019)
- 영국 개인정보보호위원회 인공지능 및 개인정보 보호에 대한 지침(2020)
- 호주 연방과학산업연구원 인공지능 8대 윤리 원칙(2019)
- 일본 내각관방 인간 중심의 인공지능 사회 구현을 위한 원칙(2019)
- 한국 과학기술정보통신부 사람이 중심이 되는 AI 윤리기준(2020)
- 방송통신위원회 이용자 중심의 지능정보사회를 위한 원칙(2019)
- 네이버 네이버 AI 윤리 준칙(2021)
- 카카오 알고리즘 윤리헌장(2018 제정, 2019·2020 개정)

법률 제정 X  
원칙과 가이드라인에 한정

한국  
: 급격한 전 분야 AI 도입에 비해  
논의가 늦은 편

을 보이는 것과 달리, 우리나라는 아직 인공지능 알고리즘 도입 때의 위험성에 대한 평가조차 이뤄지지 않은 상태인데도 행정 영역에서 서둘러 도입하려는 움직임이 있다"고 우려했다.

# 국제 표준화 현황

## IEEE 윤리지침

- 2016년, “Ethically Aligned Design”
- 최초로 AI 관련 지침을 정의했으나 업무 자동화, AWS 등의 이슈 중 하나
- 5개 쟁점: Human rights, Well-being, Accountability, Transparency, Awareness of misuse

## ISO/IEC 합동 기술 위원회(JTC)

- 2018년 4월 인공지능 표준화 전담 SC42 신설
- WG3에서 인공지능 신뢰성(Trustworthiness) 다룸

ICS > 35 > 35.020

### ISO/IEC DTR 24368

Information technology — Artificial intelligence — Overview of ethical and societal concerns

---

**GENERAL INFORMATION**

Status : Under development

Edition : 1

Technical Committee : ISO/IEC JTC 1/SC 42 Artificial Intelligence

ICS : 35.020 Information technology (IT) in general

---

**NOW**

UNDER DEVELOPMENT  
ISO/IEC DTR 24368  
Stage: 30.20

00 10 20 30 Committee 40 Enquiry 50 60 90 95

ICS

### ISO/IEC AWI 24029-2

Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods

---

**GENERAL INFORMATION**

Status : Under development

Edition : 1

Technical Committee : ISO/IEC JTC 1/SC 42 Artificial Intelligence

ICS :

---

**NOW**

UNDER DEVELOPMENT  
ISO/IEC AWI 24029-2  
Stage: 10.99

00 10 Proposal 20 30 40 50 60 90 95

# 한국 현황

- 2020년, ‘사람이 중심이 되는 AI 윤리기준’, 3대 원칙 10대 요건
  - 윤리 기준으로 활용하는 점 강조(기술 발전, 불안감 고려)
- 2021년, AI 개인정보보호 자율 점검표
  - 운영자와 개발자 대상 제작, 단계별 확인 항목 체크리스트 형태

□ (3대 기본원칙) ‘인간성을 위한 인공지능’을 위해 고려해야할 기본원칙

## ① 인간 존엄성 원칙

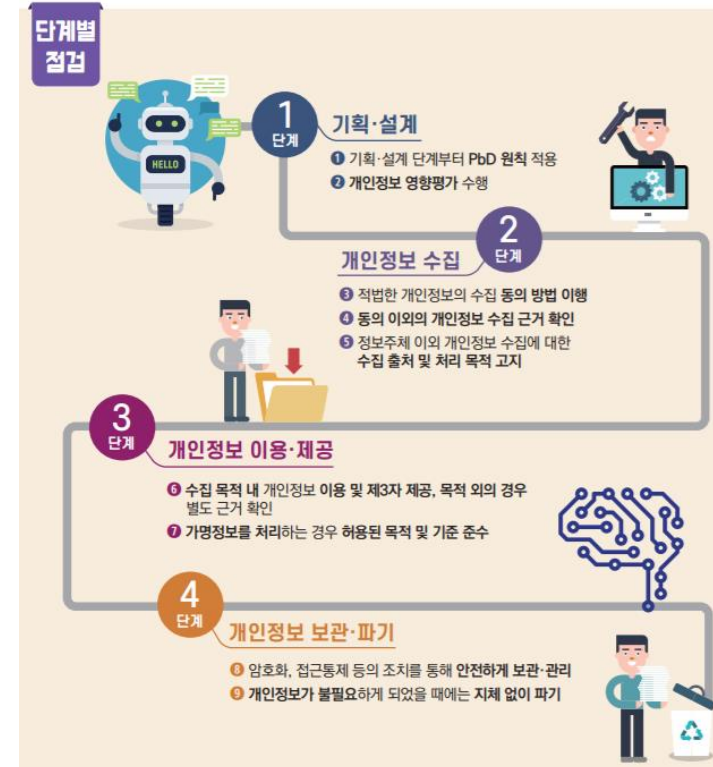
- 인간은 신체와 이상이 있는 생명체로 인공지능을 포함하여 인간을 위해 개발된 기계제품과는 교환 불가능한 가치가 있다
- 인공지능은 인간의 생명은 물론 정신적 및 신체적 건강에 해가 되지 않는 범위에서 개발 및 활용해야 한다.
- 인공지능 개발 및 활용은 안전성과 견고성을 갖추어 인간에게 해가 되지 않도록 해야 한다.

## ② 사회의 공공선 원칙

- 공동체로서 사회는 가능한 한 많은 사람의 안녕과 행복이라는 가치를 추구한다.
- 인공지능은 자동정보사회에서 소외되기 쉬운 사회적 약자와 취약 계층의 접근성을 보장하도록 개발 및 활용되어야 한다
- 공익 증진을 위한 인공지능 개발 및 활용은 사회적, 국가적, 나아가 글로벌 관점에서 인류의 보편적 복지를 향상시킬 수 있어야 한다.

## ③ 기술의 합목적성 원칙

- 인공지능 기술은 인류의 삶에 필요한 도구라는 목적과 의도에 부합되게 개발 및 활용되어야 하며 그 과정도 윤리적이어야 한다.
- 인류의 삶과 번영을 위한 인공지능 개발 및 활용을 장려하여 진흥해야 한다.



# 한계점

- 개인정보 관리에 치우친 내용
- 개발/운영자의 낮은 윤리 의식, 필요성 체감
- 관련 가이드라인의 포괄적인 단어, 낮은 효용성
- 자율 점검표 및 윤리강령 교육 정도 평가 척도의 부재

Do (실행)
<b>7장. 지원</b> ▶ 실행을 위해 조직이 지원해야 하는 것들은 아래와 같습니다. - 수립, 실행, 유지, 보수하는 전 과정에 걸쳐 시간, 인원, 예산 설비를 비롯한 자원 제공 - 역량/적격성(관련 업무를 수행하는 인원들에게 필요한 역량 결정, 고용 절차 및 징계 시 부패방지 방침 및 시스템 준수 요구) - 인식과 교육 훈련(최신화된 교육을 정기적으로 제공) - 의사소통(부패방지 방침 전체를 내·외부 관계자들에게 전달, 소통) - 위 내용 모두를 문서화된 정보로 보유
<b>8장. 운용</b> ▶ 운용 단계에서는, 부패방지경영시스템의 실행 사전과 사후, 그리고 진행과정 중에 수행해야 하는 총 열 가지의 주요 요소들에 대한 가이드라인을 제시합니다. ①운용 기획 및 관리, ②실사, ③재무적 관리, ④비재무적 관리, ⑤조직은 물론, 협력사 등을 비롯한 비즈니스 관련자와 관련된 부패리스크 관리 수단, ⑥부패방지에 대한 의지 표명, ⑦선물, 접대, 기부 및 유사한 편익, ⑧부패방지 관리의 부적절성 파악 및 관리, ⑨문제, 우려사항 제기, ⑩부패에 대한 조사 및 조치


ISO 37001 부패방지 관련 가이드라인

"우리사회 모두 AI 윤리 감수성 갖춰야"

결국 기술 보완만이 능사가 아니라 개발자는 물론 사용자 역시 'AI 윤리 감수성'을 끌어올려야 한다는 게 장 원장의 주장이다. 그는 "개발자들은 전공 지식만 배우다 보니 아무래도 윤리적 민감도가 떨어질 수 있다"며 "AI 기술을 개발할 땐 인문·사회 전공자들을 참여시키는 등의 노력이 필요하다"고 조언했다. 실제 스캐터랩이 이루다 출시 전 공개한 기술 개발 과정을 담은 30분짜리 영상을 보면 얼마나 성능이 좋아졌는지만 나열돼 있을 뿐, 정작 AI 윤리 문제 등에 고민은 전혀 담겨 있지 않다. 장 교수가 윤리적인 문제에 주목한 배경이다.

**확인 3** 개인정보 유출 사실을 알게 된 경우 지체 없이 해당 정보주체에게 관련 사실을 ☐☐☐ 알리고 필요한 조치를 하기 위한 준비가 되어 있는가? (법§34, §39의4, 영§40)

- 개인정보 유출 시 유출항목, 유출시점·경위, 유출 피해 최소화 방법, 피해 발생 시 신고 접수 방법, 피해 구제절차 등을 정보주체에게 알리고, 필요한 조치를 취해야 함

 관련 규정: 법 제34조, 제39조의4, 영 제39조, 제40조, 제48조의4, 표준 지침 제25조~제29조 등

 **참고사항**

1. 자세한 사항은 「개인정보 유출 대응 매뉴얼(20.12)」 등 참고
2. 개인정보보호 포털([www.privacy.go.kr](http://www.privacy.go.kr)) : "민원마당-개인정보유출·침해신고" 참고

AI 자율점검표 항목 14 확인 3



## 2. 주제 및 일정

### 주제

국내 인공지능 기술의 윤리적인 발전을 위한 정책, 산업, 교육 분야에서의 보완 대책 도출

### 7~8월

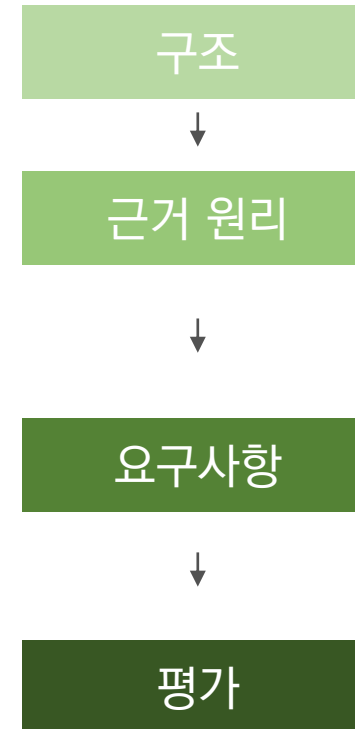
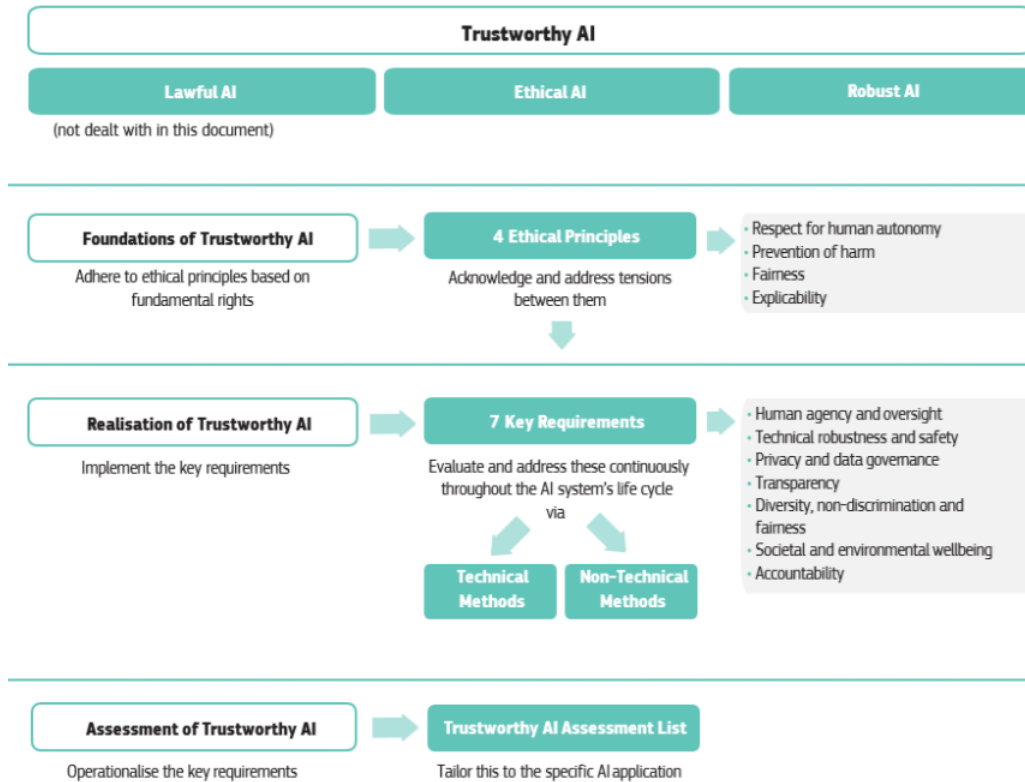
- 주요 국가 및 국내 AI 윤리 가이드라인, 지침 분석
- 주요 국가 정책, 산업, 교육 관련 인터뷰 대상 탐색

### 9월

- 인터뷰 진행
- 정책, 산업 교육 각 분야별 보완 방향 도출

# 3. 인터뷰 - 분석자료

## EU 신뢰할 만한 인공지능 가이드라인(2019)



- 7가지 조건 제시, 실질 평가의 필요성과 그 예시 제시

**Key guidance derived from Chapter II:**

- ✓ Ensure that the development, deployment and use of AI systems meets the seven key requirements for Trustworthy AI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability.
- ✓ Consider technical and non-technical methods to ensure the implementation of those requirements.

- 인간 주체성과 감독(일본, 영국), 기술 안정성(미국), 개인정보와 데이터 거버넌스(한국)
- 투명성, 다양성과 공정성, 사회 환경 복지, 책임성 – 유럽

## 인터뷰 대상

Sam Jungyun Choi (UK, Covington)

Jiyoo Chang (US, Partnership On AI)

Santeri Räsänen (Finland, University of Helsinki)

# Jiyoo Chang

- Jiyoo Chang
- Research Fellow
- San Francisco, US
- EMAIL : [jiyoo@partnershiponai.org](mailto:jiyoo@partnershiponai.org)

Jiyoo Chang is researching how transparency can be operationalized in AI systems through machine learning documentation. Previously, she developed and implemented software tools to make precision medicine more accessible to diverse populations.



**PARTNERSHIP ON AI**



# Question1



참여하고 있는 ABOUT ML에 대해 개발하면서 가장 염두에 두는 것은 무엇입니까? 현재 단계가 무엇인지 간단히 말씀해 주시겠습니까?

As for the about ML you are participating in, what do you have in mind the most in the development? Could you tell us briefly what the current stage is?

ABOUT ML 프로젝트에서 저는 AI 실무자가 문서화 작업을 할 때 직면하는 실질적인 장벽을 이해하는 것에 가장 중점을 두고 있습니다. 현재 저는 기계 학습 수명 주기에 관련된 다양한 이해 관계자를 인터뷰하고 모범 사례에 대한 인사이트를 수집하는 질적 연구를 수행하고 있습니다.

For the ABOUT ML project, I am focusing most on understanding practical barriers that AI practitioners face with implementing documentation practices. Currently I am conducting a qualitative study, interviewing diverse stakeholders who are involved in the machine learning lifecycle and gathering insights on best practices.



## Question2



이 프로젝트에서 작업할 때 참조하는 국제 표준이 있습니까? 그렇지 않다면 따라야 할 매뉴얼에 대해 알려주실 수 있습니까?

Is there an international standard that you refer to while working on this project? If not, could you tell us about the manual being followed?

몇 가지 데이터시트, 모델 카드, 팩트시트, 데이터 영양 라벨을 포함하여 이 기계 학습 문서를 통해 수행된 초기 작업을 참조합니다.

We are referring to initial works done in this space of machine learning documentation, including Datasheets, Model Cards, Factsheets, Data Nutrition Labels, to name a few.



# Question3



연구자로서 미국이 AI 윤리에 대해 얼마나 중요하다고 생각하십니까? 그런 생각이 들게 하는 케이스가 있다면 소개해 주시겠습니까?

As a researcher, how important do you feel the US now regards AI ethics? If there is a case that makes you feel that way, could you introduce it to us?

AI 윤리는 업계에서 화두가 되었고 더 많은 기업이 윤리적 AI 구축의 중요성을 인식하고 있습니다. 앞으로 몇 년 동안 미국 FTC와 같은 곳에서 공정하고 윤리적인 AI에 관한 더 많은 규정을 기대할 수 있습니다.

AI ethics has become a buzzword in the industry and more businesses are becoming aware of the importance of building ethical AI. In coming years, we can expect more regulations regarding fair and ethical AI from places like the US FTC.



## Question4



2015년 테이 사건과 비슷하게, 2021년 한국에서는 이루다 사건이 발생했습니다. 한국의 상황에 조언을 주실 수 있습니까?

A few months ago, there was an ethical controversy in South Korea with a chatbot named Iruda. The case was quite similar with Tay, AI Twitter account in 2015. The Iruda incident clearly demonstrated the level of AI ethics consideration of South Korea and more interesting to developing AI than AI ethics problem. In this situation, Is there any piece of advice to give to this situation in Korea?

이러한 문제점을 예방하기 위해서는 엄격한 AI 윤리가 필요합니다. 일반적인 AI 윤리 뿐만 아니라 강도가 높은 회사 자체 AI 윤리를 만드는 것이 도움이 될 수 있습니다. 다른 작은 기업들은 AI 윤리 분야의 다른 선두 기업들이 하는 일을 벤치마크 할 수 있고, 또한 시행 중인 규정(예: EU의 GDPR)을 이해하고, 이러한 통찰력을 한국의 사회에 적용하는 것이 도움이 될 것이라고 생각합니다.

To prevent this kind of issues, strict AI ethics are required. Not only general AI ethics, but also creating strict company's own AI ethics can be helpful. I think it'd be helpful to keep up with what other leading companies in this space of AI ethics are doing, and understanding regulations that are put in place (such as EU's GDPR), and also apply those insights in the context of Korea as a society.





# Jiyoo Chang 인터뷰 정리



가장 많은 인공지능 업체를 보유한 미국 또한 AI 윤리에 관한 관심이 많고 그 중요성을 인지하고 있습니다. 지금까지 나타났던 여러 인공지능의 윤리적 문제 재발 방지를 위해선 엄격한 AI 윤리 규정이 필요하다고 생각합니다.

현재 큰 영향력을 행사할 수 있는 대기업들은 자체적으로 윤리 원칙을 제정하고 있습니다. 해당 윤리 기준을 점차적으로 엄격화 하고, 선도적인 기업의 방식을 타 기업들이 벤치마크 하는 선순환구조는 앞으로도 AI 윤리 문제를 예방하는데 있어 중요하게 작용할 것입니다.

# Sam Jungyun Choi

- Sam Jungyun Choi
- Associate
- LONDON +44 20 7067 2054
- EMAIL : [jchoi@cov.com](mailto:jchoi@cov.com)

Sam is an associate in the technology regulatory group in the London office. Her practice focuses on European data protection law and new policies and legislation relating to innovative technologies such as artificial intelligence, online platforms, digital health products and autonomous vehicles.

COVINGTON



# Question 1



AI 관련 법률 문제가 실제로 발생합니까? 법률적 근거가 없는 상태에서 변호는 어떻게 이루어 집니까?

Do AI-related issues actually occur frequently in lawsuits? How are the defenses made even if the official legislation does not exist?

AI와 관련되었다 해서 법률적으로 크게 다른 것은 없습니다. 예를 들어 AI 공장 기계에 의해 사람이 다쳤고, AI 제조사 혹은 어떤 운영자에게 책임이 있는가를 따지게 된다면, 지금까지 AI 관련된 제품의 피해는 기존의 책임 문제의 원칙에 따라 판단되는 경향이 있습니다. 하지만 이러한 전통적인 법적 원칙이 AI 제품 및 서비스에 잘 적용되지 않을 때가 생기게 되고, 이러한 문제점을 해결하기 위해 새로운 AI법률이 필요합니다.

We don't really see a huge trend in AI related lawsuits. for example, if machinery made somebody injured, who is responsible, if it's a manufacturer or factory operator. That sort of case that around from long time so far. We've seen a trend of AI related products causing harm being adjudicated based on the principles of existing liability question, but the problem that arises is new AI laws, trying to address is around when these traditional legal principles don't apply very well to AI products and services. That's what this new law is trying to target.



# Question 1 – 추가 질문



ICO와 EU가 각각 정의하는 고위험 AI의 기준이 약간 다른 것 같습니다. 그 이유는 무엇입니까?

It seems like that the criteria for high risk AI defined by ICO and the EU are slightly different. Why is it?

ICO는 데이터 보호법을 담당하는 규제 기관입니다. EU 규칙은 제품 안전법에 더 중점을 두고 있으며, 데이터 보호법은 데이터 품질에 중점을 두고 있으며, 데이터가 올바르게 수집되었는지 확인합니다. ICO의 '고위험 AI'는 AI 제품이 **개인의 기본적 프라이버시에 대한 영향을 미칠 위험**이 높다는 것을 의미합니다. EU의 제품 안전은 **사람들에게 신체적으로 등 해를 끼칠 위험**이 높다는 의미입니다. 그래서 두가지 지침에서 고위험 AI는 다른 의미로 사용됩니다.

ICO is a regulator that's responsible for data protection law and data protection law is a different body of law altogether from product. The EU rules are focused more on products safety law, data protection law is focused on the quality of the data, Making sure that data is collected rightly. So when the ICO says, high-risk AI, they mean AI products has high risk of affecting individual, fundamental human rights to privacy. When the EU talks about product safety, they're talking about having a high risk of harming people, like physically harming people not their fundamental right to privacy, That's how the two guidelines are different.



## Question 2



AI 관련 기업과 컨설팅을 진행하면서, 기업이 가장 간과하기 쉬운 부분이 있었다면 무엇입니까?

While consulting with a company related to AI, what is the most frequently overlooked part which should be checked?

제가 알게된 가장 큰 문제점은 기업들은 기계가 원하는 모든 것을 하게 하려한다는 것입니다. 또한 기업들은 인간의 감독을 적용하는 데 충분한 투자를 하려 하지 않습니다. 인공지능은 시간이 지남에 따라 예상하기 힘든 다양한 방식으로 발전합니다. AI가 우리가 의도한 대로 작동하는지 확인하기 위해 엔지니어나 담당자가 AI를 주의 깊게 모니터링하는 것은 정말 중요합니다. 단순히 이것을 인지하는 것만으로도 AI와 관련된 많은 위험을 완화하는 데 도움이 될 수 있다고 생각하기 때문에, 저는 그것이 가장 중요하다고 생각합니다.

The biggest thing I find is that company want to let the machine do everything that it wants to do. And they don't want to invest enough resources into providing human oversight. Artificial intelligence develops over time and it's developed in different ways in unexpected ways. It's really important that there is a group of Engineers and other supervisors who monitor the AI carefully to make sure it continues to function and the way that It's intended. I think it's important and that thought alone can help to mitigate a lot of the risk association with AI.



## Question 2 – 추가 질문



AI 제품을 생산할 때, 인간의 관리감독의 범위를 설정하는 것이 큰 문제가 될 수 있다는 말입니까?

Do you mean like the actual boundary of human oversight could be the problem when they come to this AI related products

인간의 감독이 충분하지 않으면 큰 문제가 될 수 있습니다. 몇 년 전에 우리가 겪었던 테이 문제와 한국의 이 루다 문제를 보면 기계가 원하는 대로 하도록 내버려 둔 경우와 비슷한 맥락입니다. 인간, 즉 개발자가 오용한 알고리즘에 의해 실제 윤리적 문제에 직면하게 되면 문제를 해결 하기 힘들어집니다.

Having not enough human oversight can be a huge problem. And that's when you get into issues like the TEI problem that we had a few years ago and you mentioned another case study of the IRUDA, it's in a similar vein when machines are left to do whatever they want. And then human misused algorithm and the developers don't do enough to control that that's when you run into real ethical problems.



## Question 3



ICO가 데이터를 더 강조하는 것은 알겠습니다. 그런데 EU 보다 인간 감독이나 투명성 등 ICO가 더 집중하는 분야가 있다고 개인적으로 느끼는 것이 있나요?

I understand that ICOs are more focused on data things. But do you have any personal perspective that there are areas where ICOs are more focused than the EU, such as human oversight and transparency things?

여러 나라는 GDPR과 비슷한 정보 보호법을 채택하고 있습니다. GDPR보다 더 이전에 고유의 법이 있었던 한국의 정보보호법과는 다릅니다. GDPR의 기본은 투명성과 관련되어 있고 사용 방식에 대한 더 많은 권리를 사람들에게 부여하는 것입니다. AI 기술은 데이터 사용을 기반으로 하기 때문에, 가장 관심이 가는 것은 개인 정보 데이터입니다. **ICO 가이드라인은 광범위합니다.** AI를 굉장히 세부적으로 이야기하고 여러 윤리문제에 대해 이야기 합니다. 그러나 ICO는 European Commission(EC) 과는 제품 안정성과 적용에 대한 그 시각이 다릅니다.

Different countries have been adopting similar data protection. Laws to the GDPR is very different from Korea's data Protection Law because Korea Had a its own privacy law. The basis of data protection law in GDPR is all about transparency and giving people more rights over how their data is used. Because the basis of most AI technology is the use of data, most interesting things are personal data. I think ICO guidance is very far-reaching because it talks about AI in very specific terms and goes into a lot of ethics. But as I said, the focus of ICO is very different from the European commission when it try to legislate new product safety



## Question 3 – 추가 질문



ICO가 EU 가이드라인보다 더 구체적인 가이드라인과 경계가 명시되어 있는 느낌입니다.

I feel like the ICO gives more specific guidelines and boundaries than the EU guidelines, right?

ICO는 현존하는 법률의 준수에 대해 이야기합니다. GDPR은 이미 법이기 때문에 이것을 위반하면 위법입니다. 따라서 ICO는 AI 제품을 만드는 회사들에게 개인정보 보호 법을 준수하기 위해 무엇을 해야하는지 알려줍니다. 다른 ICO 가이드를 보면, 아주 구체적인 지침을 주는 것을 확인할 수 있습니다. 또 EU는 제안 단계이지 아직 법이 아닙니다. 무엇이 법으로서 제정 되어야하는지, 산업 규제는 무엇인지 결정합니다. 이렇게 모두 다른 단계에 있습니다. ICO는 현존하는 법에 대해 말하고, EU는 무엇이 법이 되어야하는지에 관련되어 있기 때문에 더 높은 레벨에서 작용한다고 할 수 있습니다.

If you look at other ICO guidance, You can see, ICO issues very specific guides. but the EU laws are proposed laws they are not yet existing law. they're talking about high-level principles and what should be law, once that becomes law, the AI regulator. if you see we're at a different stage, The ICO was talking about existing laws EU is still trying to figure out what should be law and that's why it's a lot more high-level





## Question 4



영국뿐만 아니라 다른 나라의 Covington 지부에서 다른 AI 윤리 사건이 있다면 소개 부탁드립니다.

Could you introduce any legal dispute cases related to AI ethics in other branches as well as the UK?

안면 인식 소프트웨어와 관련해서, 과거 소프트웨어 회사는 인터넷에 있는 사진을 동의없이 무작위로 가져와서 알고리즘을 훈련했습니다. 그러나 규제 기관이 재제를 하자 회사는 직원 데이터를 사용하여 알고리즘을 학습시켰고, 직원 대부분이 백인 남성이기 때문에 백인 남성에게만 높은 정확도를 보이게 된 사례가 있습니다. 미국에는 흑인이나 히스패닉에 범죄자가 많습니다. 경찰이 바디캠과 인공지능을 사용할 때 잘못된 사람을 체포하는 등의 사건들이 있었습니다. 또다른 예로 중국에서는 안면 인식을 사용하여 정치적으로 공공 서비스에 접근하지 못하도록 하기도 합니다. 우리 고객 일부는 이런 정치 체제를 지지하는 것으로 보일 수 있어 어려움을 겪고 있습니다. 또한 큰 기업의 사회적 책임과 공공 이미지 문제도 있습니다. 많은 고객들이 새로운 국제법과 EU법이 생길 때까지 법 집행 기관에 안면 인식 소프트웨어를 공급하지 않겠다고 하기도 합니다.

I want to tell you about the use of facial recognition software. Traditionally software companies used the internet and take random photos without telling them and trained their algorithm. When Regulators said, no, you can't do that, they started to train using their own employees data. Most of the employees were white men. So it would have 99% accuracy only for a white man. But in the US where a lot of the arrests are made against Black Or Hispanic. So when police use bodycam with AI, they had some cases in the US where somebody was falsely arrested and like a police brutality case. For instance in China, facial recognition is used to target people to arrest them to exclude them from accessing public services politically. some of our clients struggle with that because if they are seen in the press to be providing support to these regime. Second, it's a big corporate social responsibility and Public Image problem. A lot of our clients have sad, we are not going to supply facial recognition software to law enforcement until there's a new international law until their special law until there's EU law.



## Question 5



2015년 테이 사건과 비슷하게, 2021년 한국에서는 이루다 사건이 발생했습니다. 한국의 상황에 조언을 주실 수 있습니까?

A few months ago, there was an ethical controversy in South Korea with a chatbot named Iruda. The case was quite similar with Tay, AI Twitter account in 2015. The Iruda incident clearly demonstrated the level of AI ethics consideration of South Korea and more interesting to developing AI than AI ethics problem. In this situation, Is there any piece of advice to give to this situation in Korea?

인간의 감독 부족이 하나의 원인이며, 챗봇 개발자들은 그 작동 방식에 대해 세밀한 제어 장치를 가지고 있어야 합니다. 예를 들어, 테이 사건 이후 MS는 자체 가이드라인을 작성했습니다.

It's the lack of human oversight. The developers of this chatbot should have had very like close controls over how the chatbot works. Microsoft published on its website Guidelines for conversational AI.

Responsible bots: 10 guidelines for developers of conversational AI

November 2018



## Question 5

챗봇을 개발할 때, AI 윤리 문제를 대비할 수 있는 방법으로 기록 데이터로 훈련해야 하지만 **실시간 데이터로 훈련해서는 안됩니다.** (반유대적, 인종차별적, 성차별적 발언을 하는 실시간 유저들의 데이터는 챗봇을 오염시키기 때문입니다.)

when developing chatbots especially the bot should be trained on historical data, but not live data. (anti-semitic things, racist things and sexual things dirt out the chatbot in live data by people)

**과거 데이터에 대해 봇을 교육하는 경우 데이터를 필터하여 AI를 제어해야 합니다.** 매주 또는 매월 지속적으로 업데이트하여 AI를 관리해야 합니다. 새 실시간 데이터를 다운로드 하여 정리하고 AI에 공급하여 매번 AI에 깨끗한 데이터를 주입해야 합니다.

If you train the bot on historical data, you control it by cleaning data. You can keep the bot to be continually updated in every week, or every month. you download the new live data, clean it up, and then feed it to the bot and making sure that every time the bot, inject some clean data. it might be an important way to control for that.



# Question 6

다른 나라에 비해 영국은 여전히 AI 데이터 규제 과정을 매우 빠르게 진행하고 있고 심지어 EU 보다는 선진적입니다. 이러한 속도의 원동력은 무엇입니까?



There should be some issues with EK, EU, USA and a lot of countries related to AI but still UK had this regulation process really quickly and they actually made a law Even faster than the EU. Where do you think this velocity has come from?

1. EU와 UK는 모든 분야에서 **리더**로서의 위치를 유지하고 싶어합니다.
2. EU와 영국은 자국민 데이터를 사용하고 있는 미국 기업들에 대해 **산업 정책의 일환으로 통제권**을 행사합니다. (대부분의 AI 기업들은 미국에 기반을 두고 있다)
3. 유럽은 **2차 세계 대전과 나치 독일을** 겪으면서 개인정보를 통한 **시민통제와 홀로코스트** 역사를 겪었습니다. 이는 유럽의 개인정보 보호에 대한 관심으로 이어집니다.

The EU and the UK both want to position themselves as leaders. The EU and the UK to exert some sort of control over the US companies that are coming in and using their citizens data.(almost AI companies are based on USA). EU and the UK have a very long history of trying to protect human rights.Europe experience of World War II and Nazi Germany. Control of citizens by the state using information that led to the Holocaust. And a lot of other war crimes that became the basis of the interest in human rights



# Question 7



AI 규제가 AI의 발전을 저해한다는 우려에 대한 의견을 들어보고 싶습니다.

Please let me know your opinion on the importance of AI regulation and criticism that it hinders the development of AI.

국가는 규제 적용시 **높은 수준의 원칙**과 동시에 스타트업들의 개발에 지장이 가지 않을 수 있도록 **자유**를 부여하는 지점 사이에서 **균형**을 맞추는 것이 매우 중요합니다. 영국은 이러한 균형을 보여주는 좋은 모델입니다. 영국은 매우 강력한 데이터 보호법을 가지고 있으며 이러한 보호법은 데이터 소스와 데이터의 품질을 보호하고 데이터를 제공한 개인을 보호합니다. 동시에 스타트업의 인공지능 개발자들이 그들의 제품을 개발할 수 있도록 많은 자유를 부여합니다.

Regulators could try to balance that is to have high-level principles and Regulatory sandboxes that allows smaller companies to do what they want. Then when they become larger and become more profitable, they need to conform to Legislation. Therefore current situation in the UK is a good model because we are very strong data protection laws. So you're protecting the data sources and the quality of the data and protecting individuals, whose data is being used. But there is still a lot of leeway for AI developers to go and develop their products.



Regulatory sandboxes

smaller  
companies



Larger,  
profitable  
company

high-level principles

# Question 8



누군가가 악의적으로 AI를 사용했을 때, 그 책임은 악의를 가진 사람에게 있습니까? 아니면 그것을 걸러내지 못한 개발자에게 있습니까?

When someone abuses AI with a malicious mind, is the responsibility attributable to a person with a malicious mind or to a developer who failed to filter it?

1. AI를 악의적으로 이용하고자 한 사람에게 책임이 있으며, 개발자도 부분적으로 책임이 있습니다.
2. 이러한 문제는 AI가 아니어도 복지수당 부정 수급과 같은 기존에 존재했던 문제들이며 이미 입법화 되어 있습니다.
3. 이러한 문제들은 의사결정 시 기존 protocol이 AI에도 동일하게 적용됩니다. AI 사례와 기존 사례들의 유일한 차이점은 데이터의 유무입니다

I think the person who told the system the wrong data is responsible. If the developer does not have a safe guard or does not check the data, and they do not fulfill their obligations under the terms and conditions of the contract, the developer is also partially responsible. Unfair welfare supply and demand. And there is legislation. All of these problems that you have with an AI products exist in software exist in decisions, made by humans, according to protocol. The only difference is data.



# Question8

## 4. 의사 결정에서의 인간과 AI 의 차이점

Human : trained based on employee handbook or their own rules + principles-based

AI : trained based on input data based on historical decision + less principles-based

problems with data privacy



ICO's guidelines and data Privacy Law

5. 동일한 알고리즘 유형을 여러곳에서 활용하면 AI에서 발생하는 문제가 커질 수 있습니다. 예를 들어, 동일한 사회적 신용 평가 알고리즘을 은행과 학교 회사 등 모든 기관이 동일하게 사용한다면, 이것은 사회적 불평등을 증폭시킬 수 있습니다.

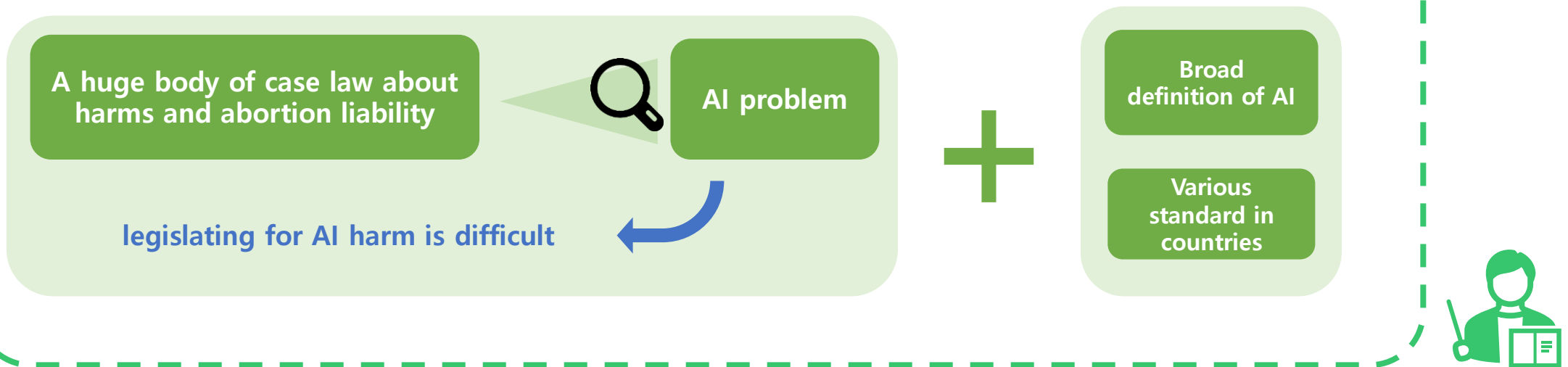
If the type of algorithm is used as a whole, there is a problem that the problem arising from AI is amplified. For example, the same social credit rating algorithm If used by all organizations of banks and school companies based on the same algorithm, this can amplify social inequalities.



# Question8

6. 3번과 관련하여, AI 문제는 기존 판례와 동일하게 적용이 되며 이는 인공지능 관련 법 제정을 더 어렵게 만듭니다. 또한, 광범위한 AI의 정의는 가이드라인 제정에 어려움을 줍니다. 이러한 상황에서, 나라별로 다양한 인공지능에 관한 법률 존재는 혼란을 야기할 수 있습니다. 그렇기 때문에 기준 지침을 만드는 것은 쉽지 않은 작업입니다.

You have a huge body of case law already trying to address those harms and abortion liability and address those problems. So AI will be viewed in that lens and I think that's why legislating for AI harm is so difficult. It can be a lot of problems for AI laws to deal with multiple countries. It already has various jurisdictions over how to deal with such issues.





# Sam Jungyun Choi 인터뷰 정리



유럽은 전세계 중에서도 AI 윤리 법률 체계에 있어서는 가장 선구적이라고 생각합니다. 개인정보 보호와 관련된 GDPR이라는 법이 있고, ICO는 이러한 개인정보 보호법을 지킬 수 있도록 회사들을 가이드 하는 역할을 합니다. EU는 인간에게 안전한 AI윤리와 관련한 입법을 통해 이를 강화하고 있습니다.

EU의 체계를 일부 벤치마킹 하여 AI 윤리에 관한 구체적인 체계를 잡는 것은 한국에 효과적인 방법이라고 생각합니다. 정부 차원에서 갱신 가능한 AI 인증제도를 도입하여 단순히 모델을 시장에 내놓는다고 끝나는 것이 아닌, 인간의 끊임없는 관리감독 및 유지보수를 점검하는 제도가 꼭 필요하다고 생각합니다.

# Santeri Räisänen

- Santeri Räisänen
- PhD student, University of Helsinki
- Finland Helsinki
- EMAIL : [santeri.raisanen@helsinki.fi](mailto:santeri.raisanen@helsinki.fi)

Santeri is a PhD student in Science and Technology Studies at the Center for Consumer Society Research. His current research revolves around the technological narratives surrounding data practices and their role in sociotechnical change. He participated in foundation of 'Ethics of AI' course for educating AI ethics problems.



# Question 1



핀란드 자체에서 AI와 관련된 법이나 제도를 만들려는 시도가 있습니까? 현재 어떤 단계인지 간단히 말씀해 주시겠습니까?

Is there any attempt to create laws or systems relevant to AI in Finland itself? Could you tell us briefly what the current stage is?

현재 관련 제도는 유럽연합(EU)를 중심으로 이루어지고 있습니다. 우선, 일반 데이터 보호 규정인 GDPR은 개인에 대한 데이터를 기업이 수집할 수 있는 방식을 규정하며 이것이 AI 시스템 구축에 영향을 미칩니다. EU가 입안한 AI 규제안은 아직 발효되지 않았지만 그 내용은 아직 협상 중입니다.

GDPR, which is the general data protection regulation regulated the way data about individuals can be gathered by companies, so in that way it also effects the building of AI systems. On the other hand, the new AI regulation drafted by the EU is yet to be in effect, but the content of that regulation is still under negotiation



## Question 2



EU/핀란드는 AI와 신기술 윤리를 얼마나 중요하게 생각합니까? 만약 당신이 그렇게 느낄 수 있는 사례가 있다면, 소개해 주시겠습니까?

As a researcher, how important do you feel the EU/Finland now regards ethics of AI/new technologies? If there is a case that makes you feel that way, could you introduce it to us?

핀란드는 규제와 시장 성장 지원 사이에서 끊임없는 갈등을 하고 있다고 생각합니다. 윤리 문제의 중요성은 자주 언급되나 상당히 좁게 정의되어 있고, 경제 성장의 관점에서 제한됩니다. 현재 기업에 유리하게 작용하는 이러한 논의들에 대해 시민 사회의 역할이 앞으로 커져야 합니다. 아쉽게도 EU의 윤리에 대한 논의가 활발함에도, 실질적인 규제 조치로 이어지지 않는 않거나 그 강도가 매우 약한 편입니다.

I feel that Finland is in a constant battle between regulation and supporting growing markets. Ethics are quite narrowly defined, any possible moral edicts are restricted in terms of the growth of the economy. Currently, a large oversight in the Finnish technology conversation is the role of civil society, which is largely sidelined in favor of companies and people as individual consumers. The consumer protections, and the proposed EU law, are surprisingly much weaker than for example China's draft algorithm regulation proposal. So I feel that while the talk of ethics is big in the EU, often the talk on ethics does not translate to any real regulatory action.



## Question 3



당신의 생각에, 새로운 AI 시스템을 개발하면서 기업이나 개발자들이 간과하기 가장 쉬운 것은 무엇이라고 말할 수 있습니까? 그것은 누구의 책임이라고 말할 수 있습니까?

In your opinion, what is the easiest thing for companies or developers to overlook while developing a new AI system? Whose responsibility can it be said to be?

기술적인 문제에만 초점을 두는 것입니다. 사업 모델 그 자체, 권력 관계, 그리고 수반되는 생산의 전체 생태계가 윤리적인가 역시 고려해야 합니다. 예를 들어, 시스템이 강제 노동(예: 데이터 라벨링)으로 구축되거나 신체적 특징(많은 AI 성격 또는 감정 감지 시스템 등)을 통해 개인을 상품화하는 경우 윤리적일 수 없습니다. 이 시스템들이 완전히 편향되지 않거나 투명할지라도, 그것이 윤리적으로 타당하다는 것을 의미하지는 않습니다.

Often the ethics conversation around AI is restricted to technical issues around the AI system itself, like bias or transparency, and these are then understood as the limited conditions of an ethical AI; that if we fix these, then everything will be ethical. What is not thought of here, is whether the business model itself, the power relations therein and the whole ecosystem of production it entails are ethical in the first place. For example, a system cannot be ethical if it is built with forced labor (in data labelling for example), or if it commodifies individuals through their physical features (like a lot of AI personality or emotion detection systems). Even if these systems are totally unbiased or transparent, it does not mean they are ethically valid.



## Question 4



2015년 테이 사건과 비슷하게, 2021년 한국에서는 이루다 사건이 발생했습니다. 한국의 상황에 조언을 주실 수 있습니까?

A few months ago, there was an ethical controversy in South Korea with a chatbot named Iruda. The case was quite similar with Tay, AI Twitter account in 2015. The Iruda incident clearly demonstrated the level of AI ethics consideration of South Korea and more interesting to developing AI than AI ethics problem. In this situation, Is there any piece of advice to give to this situation in Korea?

저는 이 사건이 일반적인 윤리가 아닌 정립되지 않은 'AI 윤리'에 초점을 맞추는 것이 얼마나 해로운지를 분명히 설명해 준다고 생각합니다. 모호하고 일반적인 AI 윤리지침이 개인정보를 보호해야한다는 기본적인 책임을 가린 것 같습니다. 따라서 문제는 시스템이 AI였다는 사실이 아니라, 어떤 목적이든 상관없이 사람들의 개인 정보가 무단으로 수집되고 사용되었다는 것입니다.

I feel this incident explicates clearly how a focus on AI ethics, instead of ethics in general, is very harmful. The fact that AI ethics guidelines are vague or of a too general level does not remove the larger and more general ethical responsibilities we have to respect people's privacy and personal information. The issue here is not really the fact that the system was AI, but that people's personal data was collected and used without permission, which is unethical regardless of whether it is used for AI or other purposes.



# Santeri Räisänen 인터뷰 정리



핀란드는 EU의 AI윤리 및 개인정보 규정을 따르고 있습니다. 하지만 아직 법제화된것은 아니기에 실질적인 규제조치는 이루어지지 않고 있습니다. 핀란드는 한국과 마찬가지로 규제가 AI 산업의 성장에 방해가 되는 것을 크게 우려하는 분위기이기 때문입니다.

그럼에도 불구하고 평등과 공정을 지향하는 핀란드의 교육에 의해 기본 윤리와 도덕성이 이미 기술직군에 강력하게 작용합니다. 현 상황에 맞추어 mooc와 같이 누구나 접근할 수 있는 강의와 관련 교양 강의를 지속적으로 개발하며 AI윤리 교육을 보조하기도 합니다. AI윤리 교육을 받은 개발자 양성과 정부의 적절한 정책이 시너지를 이룬다면 관련 규제법이 나오더라도 산업의 성장이 지체되는 부작용을 최소화할 수 있을 것입니다.

## 4. 아이디어 도출 및 결론

### 정책

- EU 및 타국 가이드라인 참고, 윤리지침 구체화
- AI 제품 개발 시 관련 인증 등 안전, 적합성 확인 절차 확립

### 산업

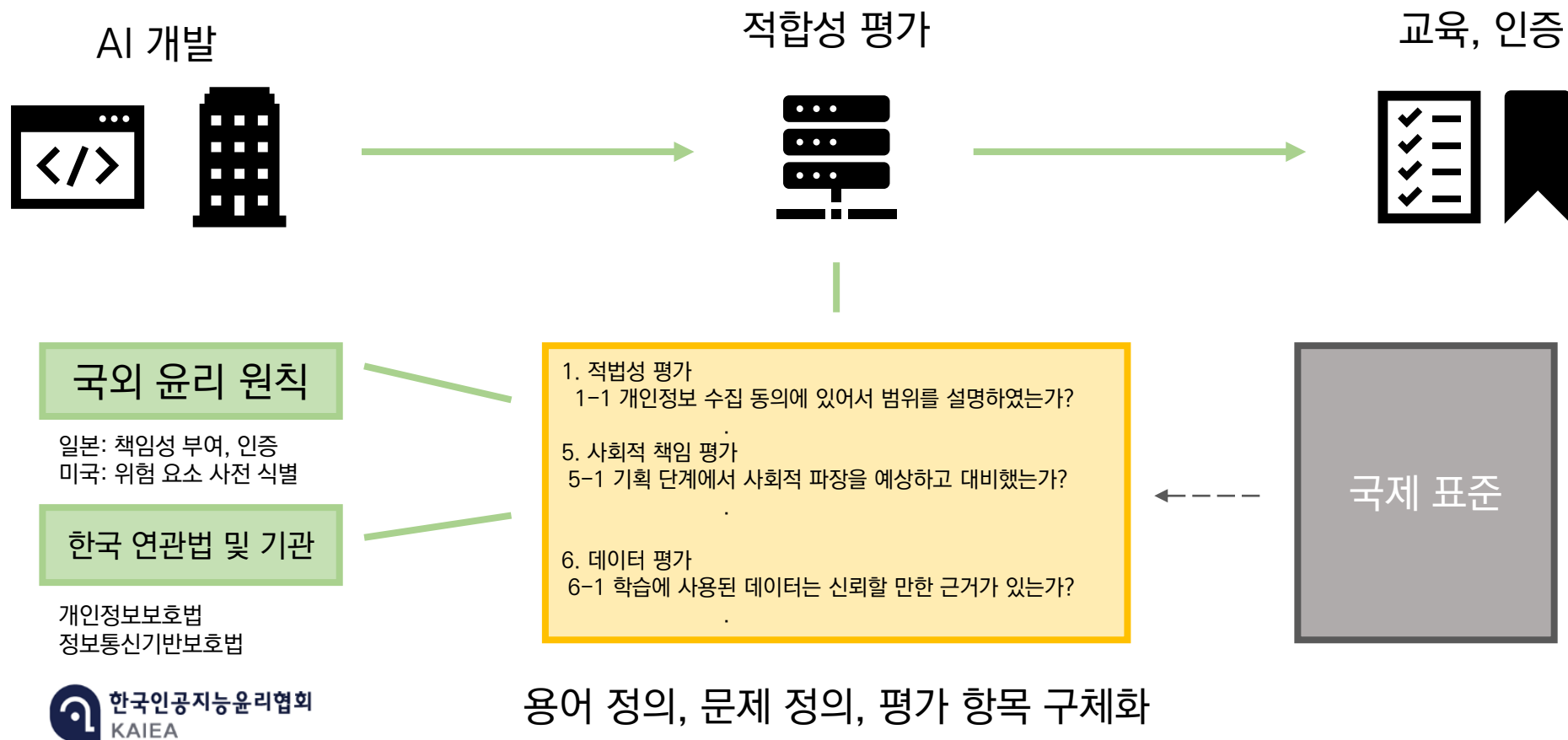
- 자체 윤리 강령 및 지침 확립
- 직원 교육 및 제품 개발에 활용

### 교육

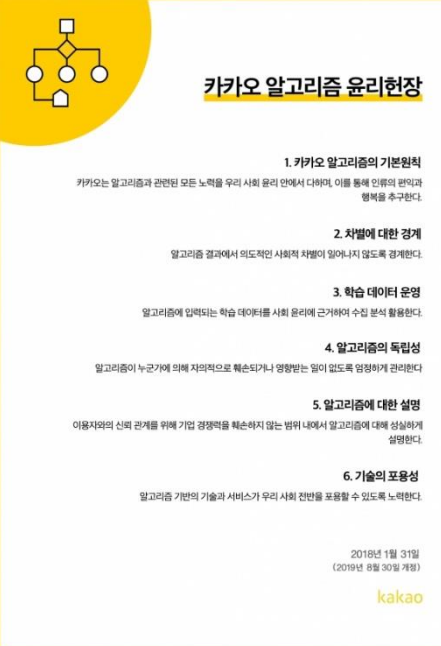
- 범분야적으로 AI와 윤리의 공존 필요성을 강조
- 개발 인력의 필수 교육을 포함, 관련 교재 개발



# 정책 분야: AI 적합성 인증



# 산업 분야: 자체 AI 윤리 준칙 확립



**카카오 알고리즘 윤리현장**

**1. 카카오 알고리즘의 기본원칙**  
카카오는 알고리즘과 관련된 모든 노력을 우리 사회 윤리 안에서 다하며, 이를 통해 인류의 편익과 행복을 추구한다.

**2. 차별에 대한 경계**  
알고리즘 결과에서 의도적인 사회적 차별이 일어나지 않도록 경계한다.

**3. 학습 데이터 운영**  
알고리즘에 입력되는 학습 데이터를 사회 윤리에 근거하여 수집 분석 활용한다.


**4. 알고리즘의 독립성**  
알고리즘이 누군가에 의해 자의적으로 훼손되거나 영향받는 일이 없도록 엄정하게 관리한다.

**5. 알고리즘에 대한 설명**  
이용자와의 신뢰 관계를 위해 기업 경쟁력을 훼손하지 않는 범위 내에서 알고리즘에 대해 상세하게 설명한다.

**6. 기술의 포용성**  
알고리즘 기반의 기술과 서비스가 우리 사회 전반을 포용할 수 있도록 노력한다.

2018년 1월 31일  
(2019년 8월 30일 개정)

kakao



**인공지능 윤리: 원칙을 넘어 실천으로**  
현장에서 논하는 AI 윤리

2021. 2. 17 (수)  
14:00 ~ 16:00

세 우리가 디지털 경제의 중요한 기반이 되는 인공지능 윤리로 주목받고 있습니다. 그러나 최근 다양한 사안은 우리에게 AI 윤리를 추상적인 원칙과 윤리라는 관점의 실천이라는 차이를 느끼게 하고 있습니다. 특히, AI 윤리라는 관점은 본질적으로 논쟁의 여지가 있는 주제입니다. 이번 행사에서는 AI 윤리라는 관점을 어떻게 이해하고 실천할 수 있는지 화제와 관련 전문가를 모시고 논의하는 시간이 될 것입니다. 관심 있는 분들의 많은 참여 부탁드립니다.

**프로그램**

2:00-2:05	원형사 그림수(서울대학교, SAP 공동연구원)
2:05-2:55	발표 세션 - 기업에서 인공지능의 AI 윤리 - 김지현(서울대학교, SAP 공동연구원) - 'AI 윤리' 실천에서 기업가치 - 송다영(삼성전자) - '내부'와 '외부'의 윤리 실천
2:55-3:00	중식
3:00-4:00	재발 세션 - 현장에서 일어나는 AI 윤리 - 사희 (인사혁신처) - 최철 (국립중앙도서관) - 김성환(서울대학교) - 박성환(서울대학교) - 송다영(삼성전자) - 권영준(삼성전자)

참가 신청  
<https://www.onofmix.com/event/222674>  
시간: 2021년 2월 17일 오후 2시 ~ 4시 (총 2시간)  
장소: 서울특별시 강남구 테헤란로 119, 11F (삼성전자빌딩 11층)  
문의: [www.onofmix.com](mailto:www.onofmix.com)

기업별 자체 AI 윤리 체계 정립  
높은 수준의 안전한 AI 개발 장려



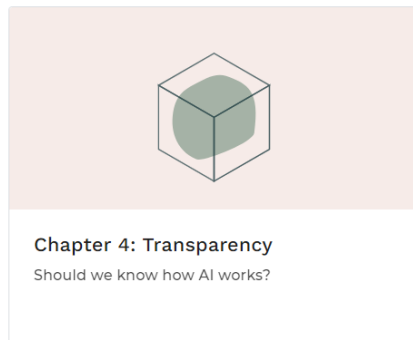
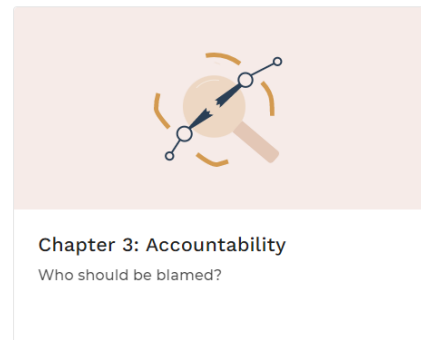
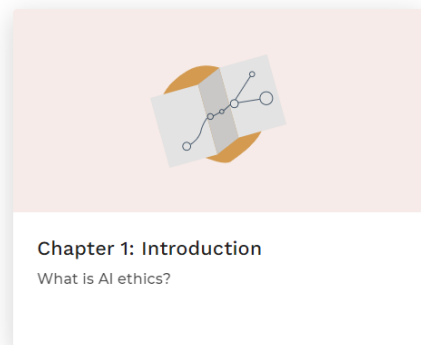
각 AI 윤리 준칙의 기업 및 스타트업 간 공유  
AI의 여러 문제들과 대응방안에 대해 협력



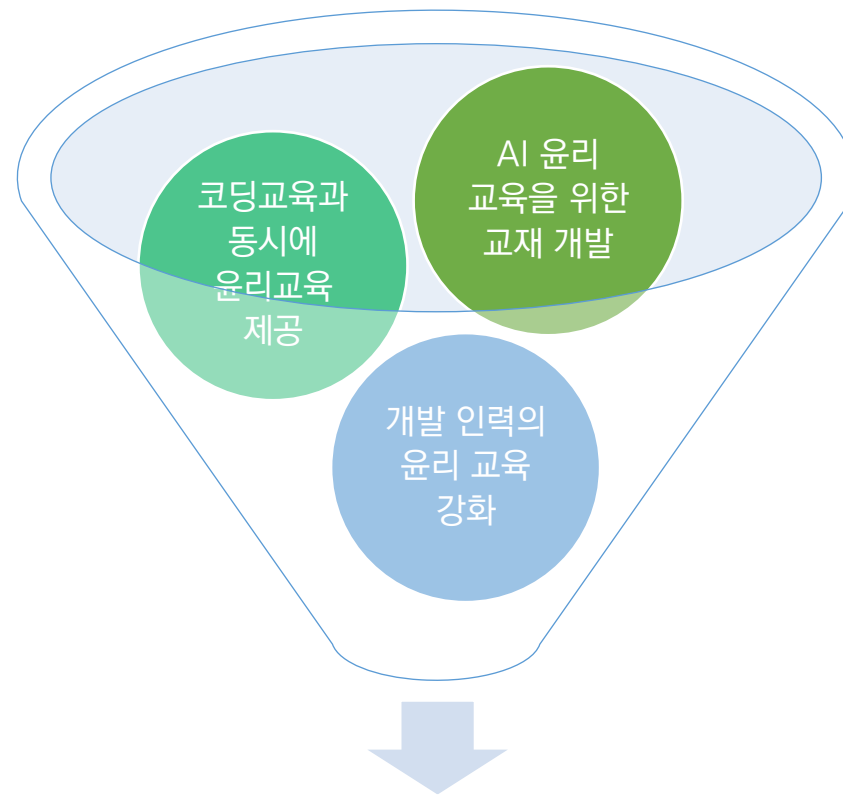
AI 산업의 긍정적 성장 기대

# 교육 분야: 개발 인력 대상 윤리 교육 이수 필수화

## Ethics of AI



[ethics-of-ai.mooc.fi](https://ethics-of-ai.mooc.fi)



윤리적인 AI 개발자 양성

# Cultural Lag?



# 감사합니다

ITM Global Challenger

Re:Union