ETRI Insight

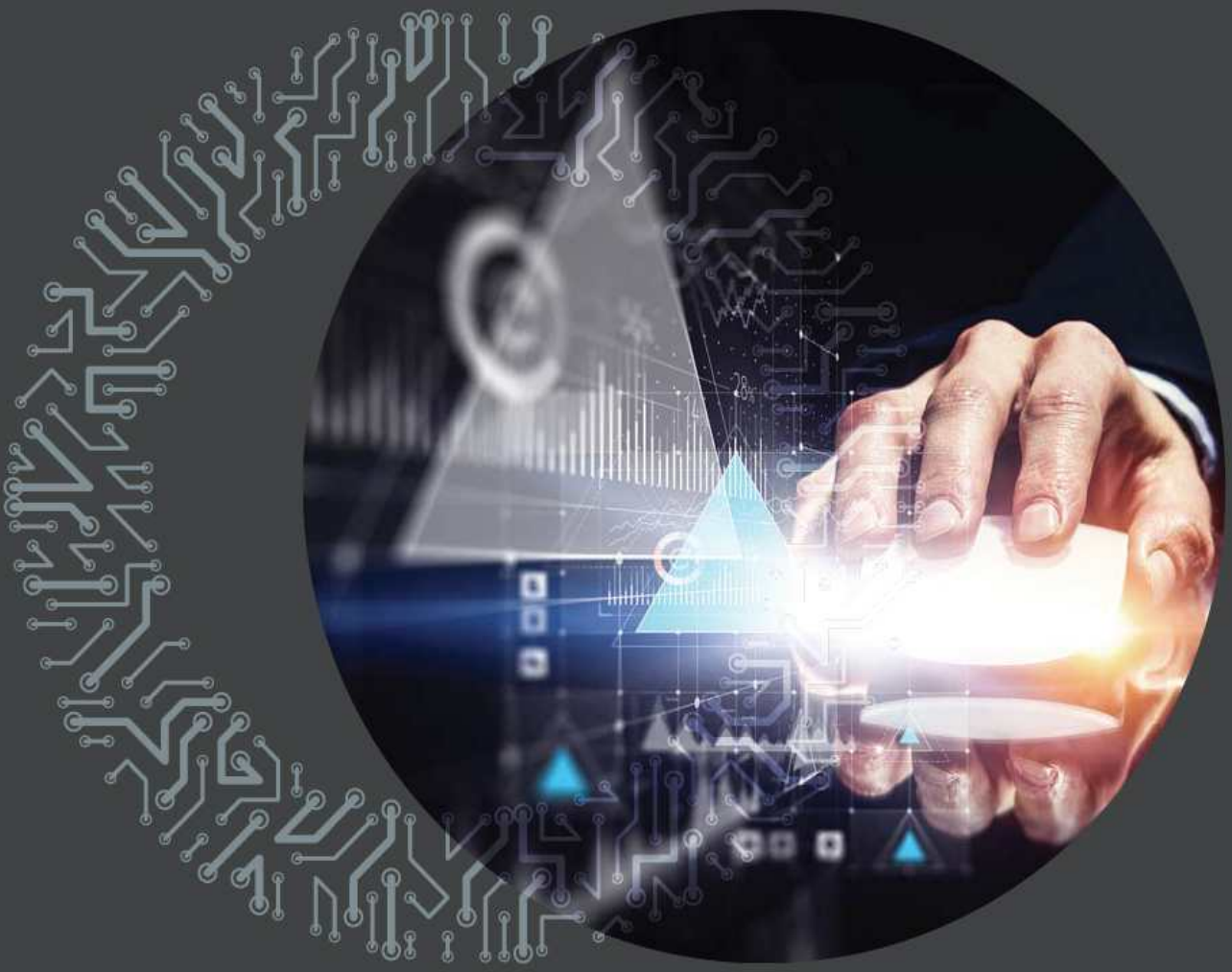
Insight Report 2017-13

기술경제연구본부 기술경제연구/산업전략연구그룹

김 문 구, 박 종 현

Insight Report

# 인공지능과 신뢰(Trust): 이슈 및 대응방안



ETRI Electronics and Telecommunications Research Institute

**※ 본 보고서의 내용은 필자의 개인적인 견해이며, 한국전자통신연구원의 공식 견해가 아님을 알려드립니다.**

CONTENTS





# 요 약


## 신뢰(Trust) 개념 및 특성

- (신뢰) 상대방의 미래 행동이 자신에게 호의적인 태도를 보이거나 최소한 악의적이지 않을 가능성에 대한 기대와 믿음으로 정의 가능
- (인공지능 시대의 신뢰 접근) 인공지능 알고리즘/설계의 기술적 복잡성 증가로 인해 예기치 않은 오류의 인지 및 충분한 설명 부족으로 신뢰 이슈 부각
  - 인공지능의 복잡한 소프트웨어 설계 → 기술복잡성 증가로 인과관계 설명 어려워짐
- (인공지능에서의 신뢰 개념) 인공지능의 활용으로 인해 인류에게 호의적인 결과발생과 그러한 발생의 과정이 투명하게 설명이 가능함으로써 예기치 않는 위험이나 위협이 줄어들 것이라는 기대와 믿음

## 인공지능 신뢰 이슈 유형

- (기술에 대한 신뢰 이슈) 인공지능 알고리즘/SW/시스템이 오류 없이 사용자가 기대한 결과가 발현될 것이라는 기대와 믿음에 대한 의문
- (제공자에 대한 신뢰 이슈) 인공지능 서비스 제공자가 악의적, 편향적인 가치관을 소유하지 않고 중립적, 객관적, 합리적 및 완결성의 서비스를 이용자가 의심 없이 사용할 수 있도록 하는 기대와 믿음에 대한 의문
- (제도/정책에 대한 신뢰 이슈) 인공지능을 안정적이고 신뢰할 수 있는 환경에서 활용이 가능하며 윤리적, 법적 책임 등의 문제로 인해 불이익을 받지 않을 기대와 믿음에 대한 의문

## 국내외 인공지능 신뢰 정책 동향

- 미국을 비롯한 해외 주요국은 인공지능의 부정적 영향, 행위의 윤리적 판단 및 법적 책임 문제 등에 대해 제도적, 정책적 검토와 사회적 합의에 의한 안정적인 인공지능 환경 구축에 집중
- 국내 또한 4차 산업혁명 및 지능정보사회에서 인공지능 신뢰와 관련 윤리, 제도, 기술개발 등 다양한 이슈에 대한 선제적 대응 마련에 자원과 역량 집중
- 궁극적으로 인공지능에 대한 신뢰, 역기능 해소를 통해 이용자에게 보다 안전한 인공지능 이용 환경 구축 요구됨

## 국내 인공지능 신뢰 인식 조사 및 시사점

- 조사 목적: 인공지능의 도입으로 인한 시스템 오류 등에 의한 우려가 예상됨에 따라 인공지능에 대한 신뢰 이슈가 부각, 이에 본 조사에서는 인공지능 신뢰측면에서 국내 일반인이 어떻게 인식하는지 설문조사 수행

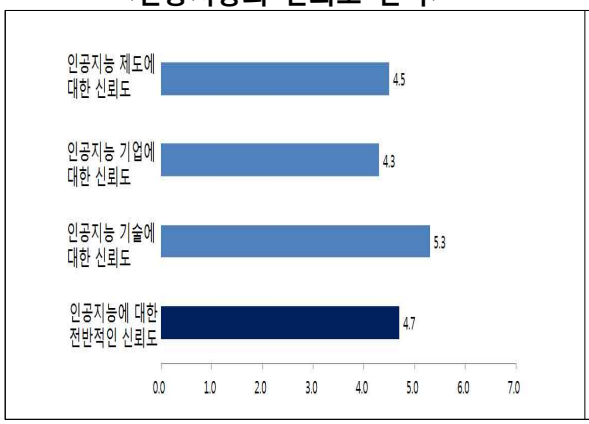
# 요 약

◐ 인공지능에 대한 신뢰도 및 위험도 인식

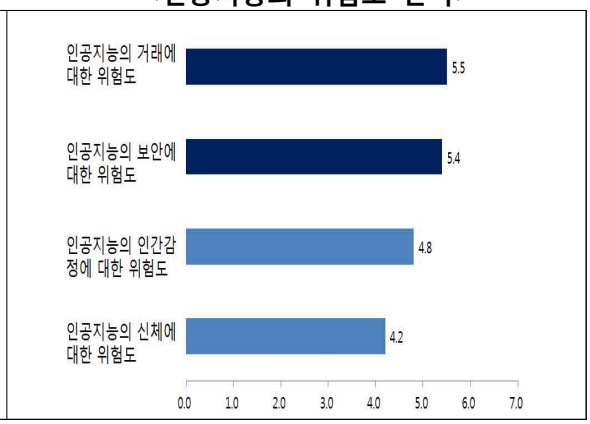
- 인공지능의 전반적인 신뢰도는 4.7점으로 약간의 신뢰를 갖는 것으로 평가되었으며, 기술에 대한 신뢰도가 가장 높게 나타남

- 인공지능이 초래할 위험에 대한 인식도는 거래에 대한 위험(5.5점)과 보안에 대한 위험(5.4점)이 가장 높게 나타남

**<인공지능의 신뢰도 인식>**




**<인공지능의 위험도 인식>**




◐ 인공지능이 이용자에게 신뢰를 제공하기 위해서는 인공지능 관련 다양한 이슈(윤리문제, 책임소재 문제, 일자일 대체, 개인정보 보호 등)에 대한 사회적 합의를 통한 합리적인 제도화 마련과 인공지능 시스템의 오류 등에 대비한 선제적 기술 개발 및 대응 필요

## 인공지능 신뢰 이슈 및 대응 방안

◐ 인공지능 기술의 도입에 따른 일자리 대체 우려, 인공지능 시스템 오류로 인한 피해 발생 우려, 개인정보 유출 및 프라이버시 침해, 악의적 인공지능 활용에 의한 각정 위협, 인공지능에 윤리적 가치판단 부여 및 법적 책임 귀속 문제 등 신뢰의 핵심 이슈로 부각

<인공지능 신뢰 핵심 이슈 및 대응방안>

| 구분 | 이슈 | 대응 방안 |
|---|---|---|
| 산업 | ■ 인공지능이 인간의 일자리를 대체할 수 있다는 우려 | ■ 인공지능으로 새롭게 창출되는 직업군에 대한 연구<br>■ 질 중심의 전문화된 일자리 창출 정책 추진 |
| 기술 | ■ 인공지능 설계시 고려하지 못한 조건의 발현, 기술의 복잡성 증가 등으로 인공지능 시스템의 오류 발생에 대한 우려 | ■ 인공지능 관련 알고리즘의 오류나 오작동의 문제를 최소화할 수 있는 기술개발 및 이를 규제하기 위한 방안 마련<br>■ 인공지능 기술에 의한 오작동을 줄이기 위한 사용자의 법적 책임 소재 명확화 |
| 데이터 | ■ 부정확한 데이터 사용으로 인한 인공지능의 피해 야기 우려<br>■ 개인정보 유출, 프라이버시 침해에 대한 우려 | ■ 해킹 방지 인공지능 서비스 개발<br>■ 데이터 처리 및 저장 방식, 사용 범위 등에 대한 규정 마련 |

# 요 약

| 구분 | 이슈 | 대응 방안 |
|---|---|---|
| 사회/문화 | ■ 악의적인 의도로 인공지능 활용에 따른 각종 위협에 대한 우려<br>■ 인공지능과의 새로운 관계 형성 의존도 증가로 비인간적, 비사회성에 대한 우려 | ■ 인공지능에의 감정이입을 이용한 인간감정의 조작을 방지하기 위한 투명한 연구개발 윤리 및 가이드라인 마련 필요<br>■ 인공지능 오남용을 사전에 탐지, 대처하는 기술 개발<br>■ 악의적인 인공지능 활용에 대비하여 인공지능 기술 및 시스템 개발 가이드라인과 중대한 위반시 법적 제재 항목의 설정 필요 |
| 윤리 | ■ 인공지능에게 인간 행위의 옳고 그름에 대한 윤리적 판단 기능 부여시 신뢰 문제 | ■ 인공지능에 윤리적 판단을 부여하는 것이 합리적인지에 대한 평가 및 사회적 합의/공론화가 선결 요건<br>■ 인공지능의 자율적 판단관련 알고리즘 개발 단계에서부터 구체적인 가이드라인 마련 필요 |
| 법적책임 | ■ 인공지능의 자율적 판단과 행동에 대한 책임 소재의 불명확성 우려 | ■ 인공지능이 책임 주체로서의 충분한 자격이 있는지에 대한 사회적 합의 필요<br>■ 인공지능 행위 결과의 사회적 수용한도에 대한 사회적 합의 필요 |

◐ 자동차 산업 사례

- 인공지능의 적용 환산으로 산업현장의 일자리 대체, 자율주행자동차 시스템의 기술적 오류로 인한 사고 발생, 운행관련 개인 데이터의 외부 유출 우려, 악의적 해킹에 의한 인적/물적 피해 우려 등의

- 자율주행자동차 사고 발생 상황시 탑승자와 보행자중 누구를 보호해야 할지에 대한 윤리적 가치판단의 신뢰 문제 발생

- 자율주행자동차의 교통사고 발생시 알고리즘 설계/개발자, 제조사, 사용자, 보험회사간의 책임 귀속 설정 문제에 대한 사회적 합의 필요

◐ 의료 산업 사례

- 인공지능의 의료진 업무 대체, 시스템적 오류로 인한 잘못된 진단 및 처방, 무검증의 불명확한 의료 데이터 사용에 따른 의료 과실 발생 및 개인 데이터 유출에 의한 사생활 침해 등의 신뢰 이슈 부각

- 인공지능이 자율적 의사결정에 의한 진단 및 처방에 대한 윤리적 가치판단 문제, 부득이한 의료사고 발생시 인공지능, 의사, 알고리즘 설계/개발자, 보험회사간 책임 소재 논란 및 사회적 공론화 필요

# Ⅰ 개념 및 특징

## 신뢰(Trust) 개념 및 특징

- (신뢰) 상대방의 미래 행동이 자신에게 호의적인 태도를 보이거나 최소한 악의적이지 않을 가능성에 대한 기대와 믿음으로 정의 가능
  - (사전적/학문적 개념) 사회의 복잡성으로 인한 두려움을 줄이기 위해 신뢰를 사용하고 상대방에 위해를 가하지 않으며 상호 신뢰할 수 있는 믿음으로 요약됨
  - (선행연구에서의 개념) 위험(Risk)의 관계, 효율적 거래관점, 효용 극대화, 협력, 대상에 대한 속성 등의 다양한 관점에서 신뢰의 개념 및 특성에 접근

**표 1 신뢰(Trust)의 다양한 개념 및 특성**

| 구분 | | 내용 |
|---|---|---|
| 사전적/학문적 개념 | 사전적 의미 | ■ 누군가 선하고 정직하다는 것이 상대방을 해하지 않고 무언가 안전하고 신뢰할 수 있다는 것을 믿는 것 |
| | 심리학적 의미 | ■ 심리학에서 신뢰는 신뢰할 수 있는 사람이 기대하는 것을 할 것이라고 믿는 것 |
| | 사회학적 의미 | ■ 사회적 복잡성을 줄이는 하나의 기제로 파악, 즉 사람들은 자신의 삶의 환경 속에 존재하는 복잡성으로 인한 두려움을 피하기 위해 이를 줄이는 방법으로서 신뢰를 사용 |
| | 경제학적 의미 | ■ 경제학에서 신뢰는 거래에서의 신뢰할 수 있는 의미로 개념화됨 |
| 선행연구에서의 신뢰 개념 | Luhmann (1988) | ■ 세계의 복잡성을 줄이는 하나의 기제<br>■ 자신의 삶의 환경 속에 존재하는 복잡성으로 인한 두려움을 피하기 위해 이를 줄이는 방법으로서 신뢰를 사용<br>■ 신뢰는 어떤 나쁜 결과로 인해서 자신의 행위를 후회하게 될지도 모르는 위험의 선택 상황에서 요구되어 짐 |
| | Misztal (1996) | ■ 경제적 교환에서 매우 효율적인 윤활유 혹은 거래를 관리하는 가장 효율적인 기제 |
| | Coleman (1990) | ■ 신뢰란 효용의 극대화를 위해 위험을 무릅쓴 의도적 행위 |
| | Gambetta (1988) | ■ 상대방에 대한 신뢰를 말한다면, 그것은 상대방이 나에게 이득이 되거나 적어도 해롭지 않은 어떤 행위를 수행할 개연성이 높기 때문에 상대방과의 어떤 형태의 협력을 시작할 것을 고려할 수 있음을 의미 |
| | 박찬웅 (1999) | ■ 한 행위자가 위험에도 불구하고 다른 행위자가 자신의 기대 혹은 이해에 맞도록 행동할 것이라는 주관적 기대<br>■ 신뢰는 사회적 관계를 전제로 그 관계 속에서 존재하며, 신뢰가 개인들간의 감시와 통제 비용을 줄일 수 있게 해주는 점에서, 협 |

| | | |
|---|---|---|
| | | 력적으로 일을 하게 해주는 관점에서 사회적 자본의 전형 |
| | 최항섭 (2007) | ■ 일반적으로 신뢰는 어떤 개인이나 집단의 미래의 행위에 대해 나의 기대에 맞게 그 행위가 이루어질 것이라고 예측할 수 있는 것을 뜻함 |
| | 윤민재 (2004) | ■ 신뢰는 진술의 진리나 사물이나 사람의 속성을 믿는 것 |

자료: Luhamann(1988), Mistzal(1996), Coleman(1990), Gambetta(1988), 박찬웅(1999), 최항섭(2007), 윤민재(2004) 참조

◐ (신뢰의 차별적 특징) 신뢰는 확신(Confidence)와의 차별적 특징을 지님

- 신뢰와 확신은 실망할지 모르는 어떤 기대와의 관련성에서 공통분모를 갖지만, 신뢰가 기대에 대한 실망 발생시 위험적 상황에서 여러 행위의 특정행위를 선택하는 특성이 있는 반면, 확신은 대안에 대한 고려 없이 기대에 대한 실망을 외부원인에서 찾으려함

**표 2 신뢰(Trust) vs. 확신(Confidence)**

| 구분 | | 내용 |
|---|---|---|
| 공통점 | | ■ 실망으로 변할지도 모르는 어떤 기대 |
| 차이점 | 신뢰 (Trust) | ■ 실망 가능성이 존재함에도 불구하고 가능한 여러 행위 가운데 특정한 행위를 선택하는 상황에서 이루어짐<br>■ 어떤 나쁜 결과로 인해서 자신의 행위를 후회하게 될지도 모르는 위험적 선택 상황에서 요구되어짐 |
| | 확신 (Confidence) | ■ 대안을 고려하지 않는 상황에서 이루어지는 것으로 확신하였다가 실망하게 되는 경우 외부적인 원인을 문제 삼음 |

자료: Luhamann(1988), 강수택(2003) 참조

◐ (신뢰의 형성) 신뢰는 주체간의 상호교류과정, 동일한 집단 소속, 제도 등의 환경적 배경에 따라 형성(최항섭, 2007; 윤민재 2004)

- 주체간의 상호교류과정: 예) 오랜 이웃관계 또는 비즈니스 거래처 사람간의 상호 지속적인 교류과정에서 신뢰 형성

- 동일한 집단 소속: 종교단체, 커뮤니티와 같은 동일집단에 속하면 상호교류가 높아지면서 신뢰형성

- 제도: 온라인 구매시 대외적 인지도 있는 사이트를 이용하는 것은 인증, 실명거래, 안전 거래 등 고객정보를 안전하게 처리하는 것에 대해 우리가 신뢰를 갖는 것임

## 인공지능에서 신뢰(Trust)의 개념

- (기존 신뢰에 대한 접근) 일반적으로 어떤 기계/시스템 등에 대한 신뢰는 우리가 기대한 기능(Function)을 목적에 맞게 충실하게 수행한다든 것을 의미함
  - 에) 로봇청소기의 명확한 기능과 작동 원리에 대한 이해와 기계 설계의 투명성에 기인한 작동의 항상성 → 기능과 결과의 인과관계에 대한 정확한 인지 가능
- (인공지능 시대의 신뢰 접근) 인공지능 알고리즘/설계의 기술적 복잡성 증가로 인해 예기치 않은 오류의 인지 및 충분한 설명 부족으로 신뢰 이슈 부각
  - 인공지능의 복잡한 소프트웨어 설계 → 기술적 복잡성 증가로 인과관계 설명이 어려워짐
  - 인공지능의 성능 향상을 위해 많은 요인/특성 고려 → 모델의 복잡도 증가 → 인간이 이해하기 어려운 결과(오류)의 발생 가능성 높아짐 → 이러한 결과에 대한 인공지능의 설명 및 인간의 충분한 이해와 신뢰 형성에 의문 증대
- (인공지능에서의 신뢰 개념) 인공지능의 활용으로 인해 인류에게 호의적인 결과발생과 그러한 발생의 과정이 투명하게 설명이 가능함으로써 예기치 않는 위험이나 위협이 줄어들 것이라는 기대와 믿음
  - 인공지능 신뢰 문제 발생 사례: 기대한 긍정적 결과와는 상이한 부정적 결과(인적, 물적 피해 등) 발생, 시스템 오류 원인에 대한 불명확성, 인공지능 행위 결과에 대한 판단의 불명확성, 데이터 유출로 인한 개인정보보호 문제 등

## 인공지능 신뢰(Trust) 이슈 유형

- (기술에 대한 신뢰 이슈) 인공지능 알고리즘/SW/시스템이 오류 없이 사용자가 기대한 결과가 발현될 것이라는 기대와 믿음에 대한 의문
  - 주요 유형 사례: 인공지능 시스템 오류 및 오작동, 복잡한 알고리즘으로 인해 결과도출에 대한 해석/설명의 어려움, 악의적 데이터에 취약한 알고리즘 등

- (제공자에 대한 신뢰 이슈) 인공지능 서비스 제공자가 악의적, 편향적인 가치관을 소유하지 않고 중립적, 객관적, 합리적 및 완결성의 서비스를 이용자가 의심 없이 사용할 수 있도록 하는 기대와 믿음에 대한 의문
  - 주요 유형 사례: 알고리즘 개발시 인간의 주관적 판단(인종차별, 성차별, 종교차별, 사회적 약자 차별 등) 반영 우려, 편향된 데이터 입력/학습에 의한 피해 발생 우려, 부정확한 데이터를 사용함으로써 야기될 수 있는 피해 발생 우려, 제공자가 악의적인 목적(범죄, 군사 등) 사용으로 인한 피해 발생 우려 등
- (제도/정책에 대한 신뢰 이슈) 인공지능을 안정적이고 신뢰할 수 있는 환경에서 활용이 가능하며 윤리적, 법적 책임 등의 문제로 인해 불이익을 받지 않을 기대와 믿음에 대한 의문
  - 주요 유형 사례: 인공지능 이용의 확산 대비 제도적 인프라 미비로 제한적 이용 우려, 인공지능의 윤리적 가치판단 주체에 대한 법제도적 미비, 교통 및 의료사고 등 인공지능으로 인한 결과에 대한 책임 주체 여부 및 인공지능과 사람간 책임 귀속 문제 등

- 신뢰는 상대방의 미래 행동이 자신에게 호의적인 태도를 보이거나 최소한 악의적이지 않을 가능성에 대한 기대와 믿음
- 인공지능 알고리즘/설계 등의 기술적 복잡도 증가, 의도/비의도 원인에 의한 부정적 결과 발생 가능성 증가
- 인공지능은 기술, 제공자, 제도/정책 등의 측면에서 다양한 신뢰 이슈 발생 가능
- 따라서, 인공지능 시대의 다양한 문제 해결을 위해 알고리즘 검증강화, 정확한 데이터 사용 및 윤리적/법적 판단을 위한 사회적 합의 등의 신뢰 형성 이슈 부각

# Ⅱ 국내외 인공지능 신뢰(Trust) 정책 동향

## 미국

- 인공지능 국가 개발 전략 계획
  - (목적) 과학기술정책국(OSTP)에서 인공지능의 도전과 기회의 명확한 이해를 목표로 인공지능의 활용을 통해 인공지능의 잠재적 혜택과 위험에 대한 준비
  - (신뢰 이슈) 인공지능의 윤리적/법적/사회적 원칙에 부합하는 인공지능 설계 개발 방법 및 구축, 인공지능 시스템이 안전하게 작동하기 위한 연구 방향 초점
- Future of Life Institute의 아실로마 인공지능 23대 원칙 공개
  - (연구 이슈) 인공지능의 부작용을 최소화하기 위한 연구 방향성에 대한 5가지 원칙

**표 3 연구 이슈 5대 원칙**

| 구분 | 내용 |
|---|---|
| 연구목표 | ■ 인공지능 연구의 목표는 방향성이 없는 지능이 아닌 유익한 지능(beneficial AI) 개발 |
| 연구비 지원 | ■ 인공지능에 대한 투자는 컴퓨터 과학, 경제, 법, 윤리 및 사회적 난제를 포함해 유익한 활용에 관련된 기금 포함 |
| 과학-정책 연계 | ■ 인공지능 연구자와 정책 입안자들 간의 건설적이며 건강한 교류 |
| 연구 문화 | ■ 인공지능 연구자와 개발자들 간의 상호 신뢰와 협력을 바탕으로 투명한 문화 형성 |
| 경쟁 회피 | ■ 인공지능 시스템 개발 조직들은 안전기준에 미달하는 것을 방지하기 위해 적극적인 협력 필요 |

자료: https://futureoflife.org/ai-principles, 양희태(2017) 참조

  - (윤리와 가치) 인간과 인공지능의 성공적인 공존을 위한 가이드라인 13대 원칙: 인공지능 시스템의 불안전한 운영, 개인 프라이버시 침해, 군사적 목적의 오용 등을 방지하기 위한 내용으로 구성

**표 4 윤리와 가치 13대 원칙**

| 구분 | 내용 |
|---|---|
| 안전 | ■ 인공지능 시스템 운영 과정상의 안전성과 보안성 보장 |
| 실태의 투명성 | ■ 인공지능 시스템이 해를 입히는 경우 그 원인 확인 가능해야함 |
| 사법적 투명성 | ■ 자동화된 시스템의 사법적 결정 참여에 대해 인간으로 구성된 감사기관의 만족스러운 설명 제공 필요 |
| 책임 | ■ 진보된 인공지능 시스템의 설계자 및 개발자들은 인공지능 시스템의 사용, 오용, 활용과 관련한 도덕적 함의의 이해관계자들이며, 이를 만들어나갈 책임과 기회가 있음 |
| 가치 연계 | ■ 고도로 자율적인 인공지능 시스템은 운영 과정상에서 그 목표와 행위가 인간의 가치와 일치되도록 설계되어야 함 |
| 인간에 대한 가치 | ■ 인공지능 시스템은 인간의 존엄성, 권리, 자유 및 문화 다양성과 양립될 수 있도록 설계되고 운영되어야 함 |
| 개인 프라이버시 | ■ 인공지능 시스템이 데이터를 분석하고 적용할 수 있는 권한을 부여받으면, 인간들 역시 자신이 생성한 데이터를 접근, 관리 및 제어 할 수 있는 권리가 있어야 함 |
| 자유와 프라이버시 | ■ 인공지능을 개인 데이터에 적용할 시 사람들의 실질적인 또는 지각된 자유가 부당하게 침해되어서는 안 됨 |
| 공유된 편익 | ■ 인공지능은 최대한 많은 사람들에게 이익을 주고 권한을 위임해야 함 |
| 공유된 번영 | ■ 인공지능으로 인해 발생한 경제적 번영은 모든 인류에게 이익이 되도록 널리 공유 되어야 함 |
| 인간 통제 | ■ 인간은 인간이 선택한 목표를 달성하기 위해 인공지능 시스템에 의사결정을 위임할지 여부와 그 방식을 선택해야 함 |
| 비 파괴 | ■ 고도화된 인공지능 시스템에 대한 통제력은 사회와 시민들이 만들어온 절차들을 파괴하지 않고 존중하고 개선시켜야 함 |
| 인공지능 무기 경쟁 | ■ 자동화된 무기와 관련한 군비 경쟁은 피해야 함 |

자료: https://futureoflife.org/ai-principles, 양희태(2017) 참조

- (장기적 이슈) 인공지능이 인간의 지성을 위협할 수 있는 수준으로 진화했을 시의 적절한 통제를 위한 5개 원칙

**표 5 장기적 이슈 5대 원칙**

| 구분 | 내용 |
|---|---|
| 역량 주의 | ■ 미래의 인공지능의 능력에 대한 합의가 이루어지지 않았기 때문에 한계치에 대한 강한가정은 피해야 함 |
| 중요성 | ■ 진보된 인공지능은 지구상의 생명체 역사에 중대한 변화를 가져올 수 있기 때문에 적절한 방법과 자원을 통해 관리되고 계획되어야 함 |
| 위험 | ■ 인공지능 시스템으로 인해 발생할 수 있는 치명적이고 실존적인 위험에 대비하는 계획 수립 및 완화 노력 필요 |
| 재귀적 자체 개선 | ■ 스스로 개선하고 복제할 수 있도록 설계되어 질적/양적으로 빠르게 확장될 수 있는 인공지능 시스템은 엄격한 안전 및 제어 조치 필요 |
| 공공의 선 | ■ 초지능은 한 국가나 조직이 아니라 광범위하게 공유된 윤리적 이념에 따라 모든 인류의 이익을 위해 개발되어야 함 |

자료: https://futureoflife.org/ai-principles, 양희태(2017) 참조

◐ 미 연방 자율주행자동차 가이드라인

- 미국 도로교통안전국(NHTSA)은 자율주행자동차 안전기준과 관련된 가이드라인 제시
- (목적) 자율주행자동차의 산업발전 촉진 및 사용자의 불안감 해소
- (내용) 안전한 도로교통 환경을 위해 연방정부의 역할과 자율주행 관련 새로운 기술 도입 및 확산을 위해 자율주행자동차 레벨 6 단계를 대상으로 15개 분야의 세부 가이드라인 마련
- (신뢰 이슈) 자율주행자동차의 탑승자 프라이버시 보호, 디지털 해킹 예방, 인공지능의 윤리적 판단 결정 기준 등

**표 6 미국의 연방 자율주행자동차 가인드라인**

| 구분 | 내용 |
|---|---|
| 데이터 기록 및 공유 | ■ 자율주행차 성능 심사를 위해 차량 결함, 성능 하락, 테스트 실패 기록을 문서화 |
| 개인정보 관리 | ■ 고객의 개인정보관리를 위해 데이터 수집의 투명성, 선택권, 보안, 책임에 관한 내용을 포함한 정책 수립 강조 |

| | |
|---|---|
| 인간과 기계 인터페이스 | ■ 인간-기계인터페이스(HMI)의 평가, 테스트, 유효성을 위해 문서화된 절차를 준용 |
| 충돌안정성 | ■ 다른 차량이 자율주행차와 충돌할 수 있는 가능성이 있기 때문에 충격 방지에 관한 기준 필요 |
| 고객교육 및 훈련 | ■ 자율주행차의 실제 도로 운행을 위한 고객교육 및 훈련을 위한 교육 프로그램 운영 필요 |
| 차량등록 및 인증 | ■ 각각의 차량 운영 모델에 따른 자율주행차 시스템의 성능과 한계를 반드시 명시 |
| 충돌이후의 반응 | ■ 충돌이후 자율주행차가 어떻게 반응했는지에 대한 평가, 테스트 사항들을 반드시 문서화 |
| 연방-주-지역법률 준수 | ■ 연방·주·지역법률을 준수하기 위한 자세한 계획과 방안들을 문서화해야 하며 자율주행차는 ODD에 기초하여 해당 지역의 도로 규제 준수 |
| 윤리문제 | ■ 자율주행차 윤리문제는 많은 사람들에게 영향을 미치기 때문에 안전, 운행, 법률 등에 따라 세세하게 고려하는 것이 중요 |
| 운행환경 관리 | ■ 제조업체 및 특정단체는 다양한 운행 환경에서의 자율주행차 평가, 테스트, 검증방법들을 정의하고 문서화 |
| 상황감지 및 대응 | ■ 자율주행 차량은 사전 충돌 상황을 감지하고 모든 사항을 고려하여 대응법을 찾아낼 수 있는 기능 필요 |
| 위험 최소화 | ■ 자율주행차 주행 중 문제가 발생하였을 때 위험을 최소화하기 위한 관련 프로세스 정리 및 문서화 필요 |
| 유효성 검증방식 | ■ 도로교통안전국 및 SAE, NIST들은 지속적으로 새롭고 혁신적인 테스트 방법을 제조업체에 적용 |

자료: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/federal_automated_vehicles_policy.pdf, NIA(2016) 참조

● IBM의 Learning to trust artificial intelligence systems

- 인공지능 시스템에 대한 남용 및 오용의 잠재성 및 책임성, 무결성에 대한 인식을 통해 인공지능에 대한 신뢰 확보 방안 제시

- 인공지능 시스템 개발시 신뢰 부여를 위해 3단계 방식 제안: (1) 인공지능 시스템 개발 전에 비즈니스 요구사항에 대한 명확한 정의와 윤리적 수용성 정립 필요, (2) 인공지능의 윤리 이슈 관련 필드 테스트를 제품 또는 서비스 배포 전에 실행, (3) 잠재적 윤리문제 이슈에 대한 지속적인 사용자 피드백을 통해 인공지능 메커니즘 설정 완료

- 인공지능 알고리즘의 책임성(Accountability): 투명성과 해석가능성이 선

결 요건

- 인공지능 알고리즘의 무결성(Integrity): 편향됨 없이 정확하고 객관적인 데이터와 인공지능 모델 구축 필요

## 일본

### 인공지능 개발 가이드라인(안)

- (목적) 인공지능 편익과 위험을 방지하기 위해 인공지능 연구개발시 유의사항 작성 → G7 또는 OECD에서 논의 자료로 활용
- (특징) 비규제적, 비구속적인 Soft Law로서 국제적으로 공유되는 지침(안)
- (신뢰 이슈) 인공지능으로 부터의 권리 이익이 침해되는 위험을 억제하기 위해 편익과 위험의 적정한 균형을 맞추기 위한 개발 원칙을 제시
  * 투명성의 원칙: 이용자와 사회적 이해와 신뢰를 얻도록 기술의 특성이나 용도에 맞춰 인고지능 시스템 입출력의 검증 가능성과 판단 결과의 설명 가능성에 유의
  * 시큐리티의 원칙: 정보의 기밀성, 완전성, 가용성 확보뿐만 아니라 인공지능 시스템의 신뢰성(의도한 대로 동작하고 권한이 없는 제3자에 의한 조작을 받지 않는 것)이나 강건성(물리적인 공격이나 사고에 대한 내성)에도 유의
  * 윤리의 원칙: 생명윤리에 관한 논의 등을 참조해 신중하게 배려, 인공지능 시스템의 학습 데이터에 포함되는 편견 등으로 인해 부당한 차별이 생기지 않게 유의

**표 7 일본의 인공지능 개발가이드라인(안)의 개발 원칙**

| 구분 | | 내용 |
|---|---|---|
| 인공지능 네트워크의 건전한 촉진과 편익증진 | 연계 | ■ 개발자는 인공지능 시스템의 상호접속성과 상호운용성에 유의 |
| 인공지능 시스템의 위험 억제 | 투명성 | ■ 개발자는 인공지능 입출력의 검증 가능성과 판단 결과의 설명 가능성에 유의 |
| | 제어 가능성 | ■ 개발자는 인공지능 시스템의 제어 가능성에 유의 |
| | 안전 | ■ 개발자는 인공지능 시스템이 엑추에이터 등을 통해 이용자와 제3자의 생명/신체/재산에 위해 |

| | | |
|---|---|---|
| | | 를 미치는 것이 없게 배려 |
| | 시큐리티 | ■ 개발자는 인공지능 시스템의 안전에 유의 |
| | 프라이버시 | ■ 개발자는 인공지능 시스템에 의해 이용자와 제3자의 프라이버시가 침해되지 않도록 배려 |
| | 윤리 | ■ 개발자는 인공지능 시스템 개발에 있어서 인간의 존엄과 개인의 자율을 존중 |
| 이용자 등의 수용성 향상 | 이용자 지원 | ■ 개발자는 인공지능 시스템이 이용자을 지원해 이용자에게 선택의 기회를 적절히 제공하는 것이 가능하도록 배려 |
| | 책임 | ■ 개발자는 이용자를 포함한 이해관계자에 대해 책임을 완수하도록 노력 |

출처: NIA(2017)

## 유럽

- (프랑스) 인공지능의 경제적/사회적 영향 전망(Anticiper les impacts economiques et sociaux de Intelligence artificilelle)
  - (목적) 인공지능을 통한 경제적/사회적 측면의 활용 및 영향 분석
  - (신뢰 이슈) 인공지능을 통한 데이터 수집 및 분석시 개인정보보호 문제 및 이슈, 인공지능 활용시 결정사항에 대한 법적/윤리적 문제 제기 및 이에 대한 책임소재 및 수용성 문제 이슈, 인공지능 개발에 따른 직업 및 업무 영역 변화* 이슈

  * 일자리 대체, 책임 면제, 사회적 격차 증가, 연구 윤리, 로봇의 법인격 등 논란 발생

- (EU) 로봇규제 가이드라인
  - (목적) 로봇공학의 적용에 의해 제기되는 윤리 및 법적 쟁점에 대한 심도 있는 분석과 유럽 및 각국의 규제당국에게 쟁점에 대한 가이드라인 제공
  - (신뢰 이슈) 법률적 검토 이슈로 ① 건강, 안전, 소비자, 환경규제 ② 법적 책임 ③ 지식재산권 ④ 개인정보보호 및 데이터 보호 ⑤ 법적 거래능력 여부

## 중국

- 국무원의 '차세대 인공지능 발전계획(Development plan for AI)
  - (목적) 전세계를 선도하기 위한 체계적인 인공지능 정책 수립 및 인공지능 강국으로 부상하기 위한 전략 마련
  - (신뢰 이슈) 인공지능이 가져올 도전과 위험성에 대한 사전예방 및 검토, 인공지능의 산업적 활용에 따른 법/제도적 제약 및 부정적 영향 감소를 위한 인공지능 관련 법률, 규정, 윤리적 기반 마련

    * 인공지능 관련 법률적 프레임 구축: 인공지능의 법률 주체 및 권리, 의무와 책임 명시

    * 인공지능 개발 활용, 보안, 관리감독, 평가체계 구축: 위험성 평가 및 강화를 위한 인공지능의 전과정을 관리/감독 체계수립, 인공지능 제품 설계 및 시스템의 복잡성/위험성/불확실성 등 잠재적 영향에 대한 체계적인 지표(데이터 남용, 개인정보 침행, 윤리적 위배 등의 행위에 대한 징계 수위 강화) 개발

## 한국

- 4찬 산업혁명 대응 계획
  - (목적) 지능화 혁명을 기반으로 경제성장과 사회문제 해결을 동시에 달성하기 위한 새로운 성장전략
  - (신뢰 이슈) 인공지능의 오작동에 대한 안전성 및 행위 결과의 책임 문제 등의 법적 이슈와 편견 없는 데이터 및 알고리즘에 의한 윤리 문제 등
- 지능정보사회 중장기 종합대책
  - (목적) 4차 산업혁명의 도래에 따라 경제/사회의 혁신적인 변화에 대응한 중장기 관점의 대응 전략
  - (신뢰 이슈) 지능정보기술의 역기능 및 오남용 최소화, 신뢰 및 안전성 인증, 인전한 데이터 사용, 오작동의 실시간 모니터링/탐지, 사이버 위협 대응 기술개발 등에 대한 추진 전략 및 R&D 투자 계획 제시

**표 8** 국내 인공지능 정책 및 신뢰 이슈

| 구분 | | 내용 |
|---|---|---|
| 4차 산업혁명 대응 계획 | 신규 법적 이슈 대응 | ■ 법제 연구와 사회적 논의: 인공지능 오작동에 대한 안정성 확보, 신기술의 결함에 대한 책임범위 명확화, 인공지능 창작물의 지식재산권 보호 필요 여부 등 |
| | 윤리 헌장 | ■ 데이터 및 인공지능 알고리즘에 사회적 편견 미반영 → 데이터 수집 및 알고리즘 개발 단계에서 기준과 절차 마련<br>- 데이터의 공정성, 신뢰성 검증, 개발자의 선량한 관리의무 등 |
| | 권리침해시 구제 방안 | ■ 인공지능의 자동화된 결정 → 결정근거에 대한 법적 근거 마련 → 피해자의 권익보호<br>- 데이터 및 알고리즘 개발자에게 반영된 편향적인 윤리적 이슈로 인한 부정적 결과 초래 가능<br>- 예) 기존 사회의 경제적 불평등 구조, 인종/성별/민족 등에 대한 사회적 편견 반영된 데이터 |
| 지능정보사회 중장기 종합대책 | 인간중심 윤리 정립 | ■ 지능정보기술의 오작동/남용 최소화 목표 |
| | 민관 합동 협의체 | ■ 지능정보사회 역기능 및 이용자 지원을 위해 기술 영향 및 위험성 등의 상시 모니터링 및 연구 |
| | 법제 정비 | ■ SW산업진흥법 개정 등: 분야별(자동차 부품, 의료기기 등) 지능정보기술의 신뢰성/안전성 인증체계 고도화 |
| | 사이버 위협 대응 지능형 자율방어체계 | ■ 사이버보안빅데이터 구축: 인공지능 기반 제품(CCTV, 자동차, 로봇 등) 및 비정형 데이터의 사이버 위협정보 수집('17년~)<br>■ 인공지능 기반 사이버 면역시스템('18년~) 및 자가방어체계 구축('20년~)<br>■ 개인 맞춤형 지능보안시스템 개발 추진(~'25까지) |
| | 지능형 자동인증기술 개발 | ■ 별도의 인증행위 없이 인공지능이 스스로 한번에 인증 및 이상 징후 발견시 대응시스템과의 실시간 연계 구축('20년~) |
| | 지능정보 SW 안정성 평가체계 마련 | ■ 인공지능 학습시 안전하고 적합한 데이터 사용 여부 및 오작동의 신속탐지/대응 여부 등에 대한 인증 방안 연구 |

출처: 과기정통부(2017), 관계부처합동(2017)

- 미국을 비롯한 해외 주요국은 인공지능의 부정적 영향, 행위의 윤리적 판단 및 법적 책임 문제 등에 대해 제도적, 정책적 검토와 사회적 합의에 의한 안정적인 인공지능 환경 구축에 집중
- 국내 또한 4차 산업혁명 및 지능정보사회에서 인공지능 신뢰와 관련 윤리, 제도, 기술개발 등 다양한 이슈에 대한 선제적 대응 마련에 자원과 역량 집중
- 궁극적으로 인공지능에 대한 신뢰, 역기능 해소를 통해 이용자에게 보다 안전한 인공지능 이용 환경 구축 요구됨

# Ⅲ 국내 인공지능 신뢰 인식조사 및 시사점

## 조사개요

◐ 조사 목적

- 인공지능의 부상으로 인공지능을 적용한 다양한 기기/제품/서비스가 일상생활에서도 활용될 가능성이 점차 커짐
- 인공지능의 도입으로 인한 시스템 오류 등에 의한 우려가 예상됨에 따라 인공지능에 대한 신뢰 이슈가 부각
- 이에 본 조사에서는 신뢰측면에서 국내 일반인이 인공지능을 어떻게 인식하는지 설문조사 수행

◐ 조사 설계

- 본 조사는 2017년 하반기에 온라인 조사를 실시하였으며, 세부 내용은 아래 표와 같음

**표 9 조사 설계**

| 구분 | 내용 |
|---|---|
| 조사대상 | ■ 전국 20~50세 성인 남녀 |
| 조사 표본수 | ■ 850명 |
| 조사방법 | ■ 온라인 조사 |
| 조사기간 | ■ 2017년 10월~11월 |

◐ 조사 내용

- 응답자 일반현황: 성별, 연령, 직업
- 인공지능에 대한 일반인의 인식: 신뢰도, 위험
- 인공지능의 기여: 개인 및 기업에 대한 기여도
- 인공지능의 신뢰 형성 방안: 협업, 법제도, 윤리, 기술개발, 글로벌 협약 등

● 응답자 현황

- 전체 응답자: 850명
- 성별: 남성 51.2%(435명), 여성 48.8%(415명),
- 연령: 20대 22.6%(192명), 30대 26.5%(225명), 40대 29.6%(252명), 50대 21.3%(181명),
- 직업군: 학생 12.4%(105명), 직장인 31.9%(271명), 자영업 15.5%(132명), 전문직 8.4%(71명), 전업주부 25.6%(218명)

**표 10 응답자 현황**

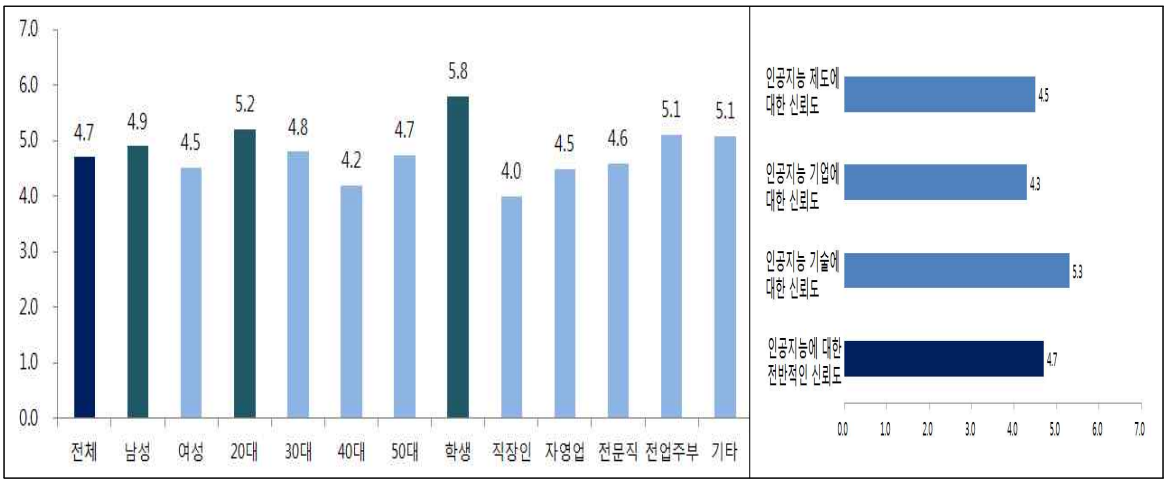
| 구분 | | 빈도 수(명) | 비율(%) |
|---|---|---|---|
| 총계 | | 850 | 100.0 |
| 성별 | 남성 | 435 | 51.2 |
| | 여성 | 415 | 48.8 |
| 연령 | 20대 | 192 | 22.6 |
| | 30대 | 225 | 26.5 |
| | 40대 | 252 | 29.6 |
| | 50대 | 181 | 21.3 |
| 직업 | 학생 | 105 | 12.4 |
| | 직장인 | 271 | 31.9 |
| | 자영업 | 132 | 15.5 |
| | 전문직 | 71 | 8.4 |
| | 전업주부 | 218 | 25.6 |
| | 기타 | 53 | 6.2 |

## 조사결과 분석

### 인공지능에 대한 신뢰도

- 인공지능에 대한 전반적인 신뢰도에 대한 조사 결과 4.7점으로 약간의 신뢰를 갖고 있는 것으로 평가
- 성별로는 남성(4.9점)이 여성(4.5점) 보다 조금 더 신뢰하는 것으로 나타났으며, 연령대별로는 20대가 5.2점으로 가장 높은 신뢰도를 보이는 것으로 나타났고, 40대가 4.2점으로 가장 낮은 신뢰도를 보임
- 직업별로는 학생이 5.8점으로 가장 높은 신뢰를 보이는 것으로 나타났고, 그 다음으로 전업주부 5.1점으로 나타났으며 직장인이 4.0점으로 가장 낮은 신뢰도를 보임
- 인공지능 관련 분야의 경우, 기술에 대한 신뢰도가 5.3점으로 가장 높게 나타났으며, 제도(4.5점)와 기업(4.3점)에 대한 신뢰도는 평균이하의 낮은 신뢰도를 보임
- 특히, 여성, 전업주부, 자영업자는 타 분야 대비 기업에 대한 신뢰도가, 남성, 40대, 직장인 그룹은 제도에 대한 신뢰도가 낮게 나타남
- 인공지능 기술 대비 기업과 제도에 대한 낮은 신뢰도가 나타남에 따라 이용자가 개인정보를 안전하고 편리하게 기업의 제품 및 서비스를 이용할 수 있는 여건 마련 그리고 제도적 측면에서 법적 책임소재 문제, 윤리 문제 등에 대한 사회적 합의를 통한 선제적 법제도 시스템 구축 필요

**인공지능에 대한 신뢰도**

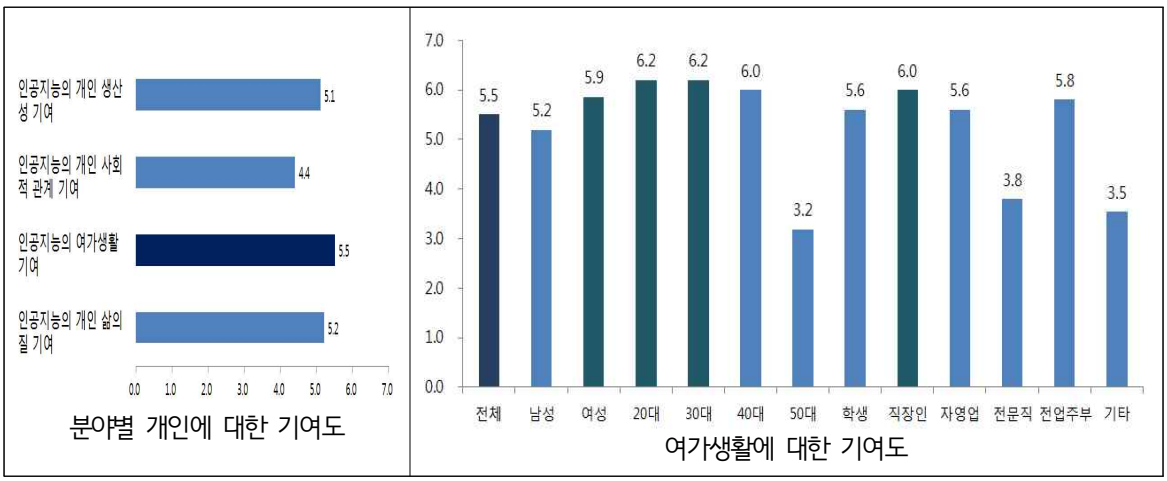



(7점 척도, 1점=전혀 그렇지 않다, 4점=보통, 7점=매우 그렇다)

## 인공지능의 개인에 대한 기여도

- 인공지능의 개인에 대한 기여도 조사 결과 여가생활 기여가 5.5점으로 가장 높은 것으로 평가되었으며, 그 다음으로 개인의 생산성 기여가 5.2점으로 나타났으며 개인의 사회적 관계 기여가 4.4점으로 가장 낮게 나타남
- 남성은 개인 생산성 기여(5.3점)를 여성은 여가생활 기여(6.2점)에 대한 선호가 높게 나타났으며, 연령층이 젊을수록 여가생활 기여를 반대로 연령층이 높을수록 사회적 관계와 삶의 질에 대한 기여를, 학생은 개인 생산성(5.7점), 전업주부는 여가생활(5.8점)과 사회적 관계(5.8점)에 대한 기여가 높을 것으로 나타남
- 가장 높게 나타난 개인 여가생활 기여도의 경우, 남성(5.2점)보다 여성(5.9점, 20~30대(6.2점), 직장인(6.0점)에서 상대적으로 높게 나타남
- 개인의 여가생활에 특화된 인공지능 활용 기술/제품/서비스 개발이 요구되며, 가정에서 활용 가능하며 주변 사람들과의 사회적 관계를 향상 내지 개선시켜줄 수 있는 인공지능을 높게 평가함

**인공지능의 개인에 대한 기여도**

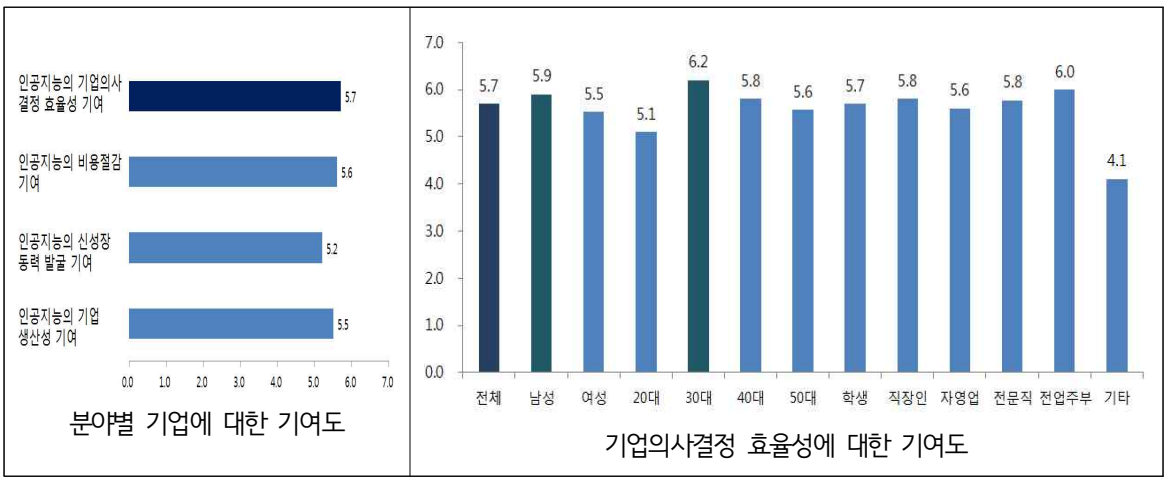



(7점 척도, 1점=전혀 그렇지 않다, 4점=보통, 7점=매우 그렇다)

## 인공지능의 기업에 대한 기여도

- 인공지능의 기업에 대한 기여도 조사 결과 기업 의사결정 효율성에 대한 기여가 5.7점으로 가장 높은 것으로 평가되었으며, 그 다음으로 비용절감 기여가 5.6점으로 나타났으며 신성장 동력 발굴 기여는 5.2점으로 상대적으로 가장 낮게 나타남
- 남성과 여성 모두 기업 의사결정 효율성 기여에 대한 선호가 높게 나타났으며, 20대는 생산성 기여, 40대는 비용절감 기여를, 직장인과 자영업은 비용절감과 의사결정 효율성에 대한 기여가 높을 것으로 기대함
- 인공지능의 기업에 대한 기여도 중 가장 높게 나타난 의사결정 효율성에 대한 기여도의 경우, 여성(5.5점)보다 남성(5.9점), 30대(6.2점)에서 상대적으로 높게 나타남
- 따라서, 향후 기업 의사결정에서 인공지능과 빅데이터 분석에 의한 자료의 보조적 활용에 대한 수요가 증가할 것으로 전망되며, 아울러 비용절감을 통한 기업의 수익성 확대에 대한 니즈가 높을 것으로 보임

**인공지능의 기업에 대한 기여도**

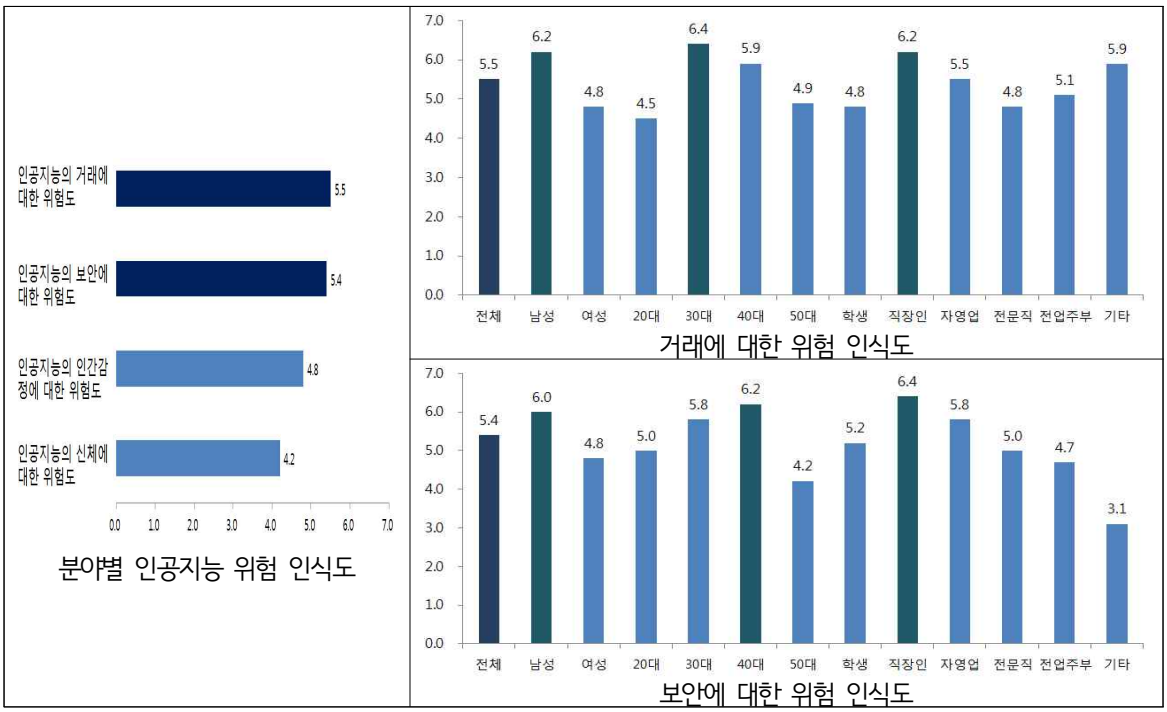



(7점 척도, 1점=전혀 그렇지 않다, 4점=보통, 7점=매우 그렇다)

## 인공지능에 대한 위험도 인식

- 인공지능이 초래할 위험에 대한 인식도 조사 결과 거래에 대한 위험이 5.5점으로 가장 높은 것으로 평가되었으며, 그 다음으로 보안에 대한 위험이 5.4점으로 나타났으며, 신체에 대한 위험은 4.2점으로 상대적으로 가장 낮게 나타남
- 대체로 거래와 보안에 대한 위험이 높게 나타났지만, 여성, 학생, 전업주부는 상대적으로 인간의 감정에 대한 위험을 높게 인식하고 있음
- 인공지능에 대한 위험 인식 중 가장 높게 나타난 거래에 대한 위험인식의 경우, 여성(4.8점)보다 남성(6.2점), 30대(6.4점), 직장인(6.2점)에서 상대적으로 높게 나타났으며, 보안에 대한 위험 인식도 유사한 결과를 보임
- 따라서, 인공지능의 오류, 해킹 등에 의한 상거래, 금융거래 등에서의 피해와 개인정보 유출 및 사생활 침해 등을 높게 인식함에 따라서 향후 인공지능의 산업적 활용시 안전한 거래 시스템 및 개인정보 보호에 대한 기술적 법제도적 대응안 마련이 요구됨

**인공지능의 위험 인식도**

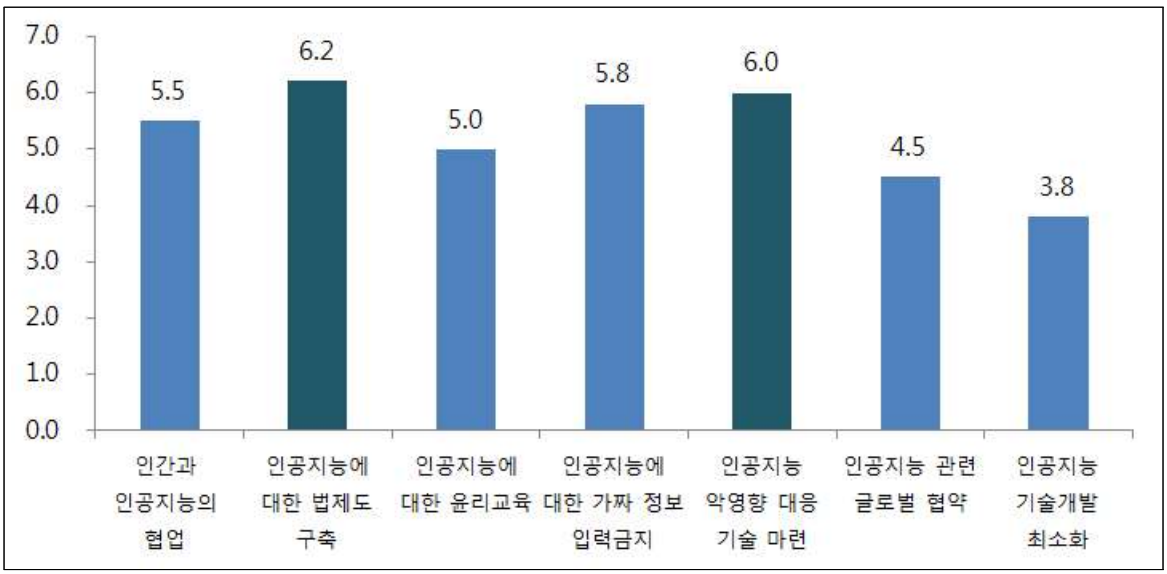



(7점 척도, 1점=전혀 그렇지 않다, 4점=보통, 7점=매우 그렇다)

## 인공지능에 대한 신뢰를 형성하기 위한 방안

- 인공지능에 대한 신뢰를 형성하기 위한 방안에 대한 조사 결과 법제도 구축이 6.2점으로 가장 높은 신뢰 형성 방안으로 평가되었으며, 그 다음으로 인공지능 악영향 대응 기술마련이 6.0점, 인공지능에 대한 가짜 정보 입력 금지가 5.8점으로 나타났고 인공지능 기술개발 최소화는 3.8점으로 가장 낮은 것으로 나타남
- 인공지능에 대한 법제도 구축의 경우 남성 보다 여성이 타 연령 대비 40대가 그리고 자영업자와 전업주부가 상대적으로 높은 점수를 보임
- 인공지능에 대한 악영향 대응 기술 마련의 경 여성보다 남성, 20~ 30대, 자영업자가 상대적으로 높은 점수가 나타남
- 인공지능이 일반인에게 신뢰성 있는 기술로 받아들여지기 위해서는 무엇보다도 법과 제도에 대한 정립과 인공지능의 부정적 영향을 최소화 시킬 기술개발이 우선적으로 해결되어야 할 과제로써 향후 인공지능에 관한 정부 정책 및 R&D의 방향성에 시사 하는 바가 클 것으로 보임

**인공지능에 대한 신뢰 형성 방안**




(7점 척도, 1점=전혀 그렇지 않다, 4점=보통, 7점=매우 그렇다)

- 인공지능에 대한 신뢰도가 어느 정도 있는 것으로 나타났으나, 인공지능이 초래할 위험에 대한 인식이 상대적으로 높게 나타남에 따라 인공지능의 산업적 활용이 확대되는 현 추세에서 인공지능에 의한 부정적 요인의 조기 해결 방안 마련이 시급히 요구됨
- 인공지능이 이용자에게 신뢰를 제공하기 위해서는 인공지능 관련 다양한 이슈(윤리문제, 책임소재 문제, 일자일 대체, 개인정보 보호 등)에 대한 사회적 합의를 통한 합리적인 제도화 마련과 인공지능 시스템의 오류 등에 대비한 선제적 기술 개발 및 대응 필요

# Ⅳ 인공지능 신뢰 이슈 및 대응 방안

## 인공지능 도입에 따른 신뢰 이슈 및 대응 방안

● (산업) 인공지능의 도입에 따른 일자리 대체 우려

- 단순 매뉴얼에 기반한 반복적인 많은 업무뿐만 아니라 지적영역에 속하는 일부 전문직 서비스 업종도 인공지능으로의 대체로 인한 일자리 감소가 빠르게 진행될 전망

  * 미래학자 토마스 프레이는 로봇과 인공지능의 발전으로 2030년이면 일자리 20억 개가 사라질 것으로 전망(중알일보, 2015)

  * 전문직 서비스에서도 로봇변호사, 로봇기자 등의 인공지능 서비스 도입으로 일자리 대체 증가 전망

  * 인공지능 기사 대비 인간기자의 경쟁력 우위 창출 중요성 증대: 인간기자가 쓴 기사에 대한 신뢰가 떨러지면 대중은 인공지능이 작성한 기사를 대신 읽을 가능성 높음 → 단순 팩트보다는 더 나은 가치 판단의 방향성 제시 필요(김대원, 2017)

**[대응 방안]**

➲ 인공지능으로 새롭게 창출되는 직업군에 대한 연구, 직업 훈련 프로그램 개발 및 교육 강화

➲ 양 중심의 일자리에서 질 중심의 전문화된 일자리 창출 정책 추진

● (기술) 인공지능 SW의 잘못된 설계, 기술의 복잡도 증가 등에 따른 오류 발생으로 인한 피해 우려

- 인공지능 시스템 설계시 고려하지 못했던 사항으로 인한 오류/오작동으로 인한 인명/재산 피해 발생 우려

- 인공지능 시스템의 기술복잡도가 증가함에 따라 오류/오작동 원인 파악이 어려워져 피해 발생에 따른 대응방안 마련이 어려워짐

  * 인공지능의 오류로 인한 인명 피해 우려가 48.6%로 나타남(SKKU 위험커뮤니케이션 연구단, 2017)

  * 자율주행자동차, 철도 등의 사람을 운송하는 교통수단의 경우 인공지능 시스템 오작동 발생시 심각한 인명 및 재산 피해 발생 가능

* 인공지능 시스템 오류 사례: 미국 쇼핑몰에서 순찰중인 인공지능 로봇의 아이 공격, 무인기에 탑재된 인공지능 프로그램의 오작동으로 결혼식장을 향하는 차량 공격 등(The ScienceTimes, 2017)

**[대응 방안]**

➲ 인공지능 관련 알고리즘의 오류 및 오작동의 문제를 방지하거나 최소화할 수 있는 기술개발 강화 및 이를 규제하기 위한 방안 마련

➲ 인공지능 기술에 의한 오작동을 줄이기 위한 설계/개발자 및 사용자의 법적 책임 소재 명확화

◐ (데이터) 부정확한 데이터 사용 및 개인정보 유출 등으로 인한 프라이버시 침해 우려

- 부정확하며 잘못된 정보에 기반한 인공지능이 생성한 결과에 대한 신뢰성을 갖기 어려우며, 그로 인해 편향 및 왜곡된 문제 야기

- 인공지능은 개인에 대한 실시간의 다양한 형태의 라이프로그 데이터(차량이용 정보, 의료정보, 대화정보 등)를 사용함에 따라 개인정보 유출시 매우 심각한 프라이버시 침해 상황에 직면 가능

* 미국 플로리다주 브로워드 카운티의 잘못된 정보에 의한 범죄 재발가능성 예측 오류 사례: 약 18,000명의 범죄자를 대상으로 향후 2년 동안 새로운 범죄를 일으킬 가능성을 인공지능을 활용한 범죄자 예측 알고리즘을 통해 분석한 결과 흑인이 백인보다 범죄 재발 가능성이 약 45% 높은 것으로 파악되었으나, 동일기간 실제 데이터를 분석한 결과 오히려 백인의 재범 비율이 높게 나타남(The ScienceTimes, 2017)

* 스마트홈 서비스, 홈내 인공지능 스피커 확산, 스마트 그리드 등의 진전으로 사적 영역에 대한 실시간의 무제한적 데이터 접근 용이성 증가 → 프라이버시 침해 가능성 증대

* 2011년 16개 공공기관의 31개 공공DB 품질조사 결과 지표별(완전성, 일관성, 유효성) 평균 오류율이 5.19%로 나타남(한국데이터베이스진흥원, 2011)

**[대응 방안]**

➲ 해킹 방지를 위한 인공지능 서비스 및 기술 개발 강화

➲ 정확한 데이터의 인공지능 적용을 위해 데이터 처리 및 저장 방식, 사용 범위 등에 대한 규정 마련

◐ (사회/문화) 악의적인 의도로 인공지능 활용에 대한 위협과 인공지능과의 새로운 관계 형성에 따른 비인간적, 비사회성 인격 형성에 대한 우려

- 인공지능 알고리즘 설계/개발자가 종교, 인종차별, 범죄 등의 악의적인 의도를 가지고 인공지능을 이용할 경우 심각한 사회문제 야기

- 인공지능에 대한 의존도가 높아짐에 따라 사람과의 커뮤니티 형성/참여를 지양, 비사회적인 성향으로 변화될 가능성 높음

* 국제 테러리스트에 의한 인공지능 활용 범죄 증가 가능성, 특정 종교 및 인종에 편향적인 인공지능 알고리즘을 설계하여 종교/인종적 부당한 차별 가능성 증대

**[대응 방안]**

➲ 인공지능에의 감정이입을 이용한 인간감정의 조작을 방지하기 위해 투명한 연구개발 윤리 및 가이드라인 마련 필요

➲ 인공지능의 오남용을 사전에 탐지 및 대응 가능한 기술 개발

➲ 악의적인 인공지능 활용에 대비하여 인공지능 기술 및 시스템 개발 가이드라인과 중대한 위반시 법적 제재 항목의 설정 필요

◐ (윤리) 인공지능의 윤리적 가치판단 부여로 인한 신뢰성 문제

- 인공지능에게 가치판단이 들어간 윤리적 판단을 부여해야 하는 것이 옳은 것인지? 만약 윤리적 판단을 인공지능에게 허용할 경우 초래될 부작용 내지 위험은 어떻게 발생할 것인지에 대한 논란 야기

- 인공지능 윤리문제는 크게 ① 인공지능 알고리즘 설계 단계에서 적용되는 윤리, ② 인공지능의 자율적 의사결정에 의한 행동에 대한 윤리 등이 핵심 쟁점

- 인간의 윤리와 부합하는 윤리적 기준을 갖지 못할 경우, 인공지능 판단이 인공지능 설계자의 윤리적 편향이 들어갈 경우, 부정확하고 편향적인 비윤리적 데이터에 기반을 둔 인공지능 학습이 이뤄질 경우 등에 대한 우려 발생

**[대응 방안]**

➲ 인공지능에 윤리적 판단을 부여하는 것이 합리적인지에 대한 평가 및 사회적 합의/공론화가 선결 요건

➲ 인공지능의 자율적 판단과 관련하여 알고리즘 개발 단계에서부터 구체적인 가이드라인 마련 필요

● (법적 책임) 인공지능의 자율적 판단과 행동에 대한 책임소재의 불명확성 우려

- 인공지능 행위로 인해 야기된 손실에 대해 인간의 개입이 없는 인공지능 스스로의 자율적 의사결정에 의한 행위에 관한 책임 배분 문제 발생

- 인공지능이 도덕적 주체가 될 수 있는지? 인공지능의 범죄 행위에 대해 인간처럼 형사책임을 물을 수 있는지?에 대한 논의 증가

- 인공지능이 야기한 결과에 대한 책임 귀속의 문제에 대한 다양한 논란 발생: (1) 예측가능성과 회피가능성에 따라 ① 인공지능의 자율적 행동에 의한 불법적인 행동에 대해 사용자가 예측하거가 회피할 수 없으므로 인공지능 사용자에게 책임 귀속 불가, ② 인공지능 사용자는 인공지능 소유자로 인공지능의 행동으로 발생한 모든 피해에 무과실 책임 귀속(임석순, 2016), (2) 인공지능의 예측할 수 없는 위험의 사회적 용인 여부에 따른 책임 귀속(윤지영, 2015)

- 제조물 책임 부여 여부 논란: 인공지능이 제조물에 해당하는지? (1) 형법적 제조물 책임을 부여하기 위해 무과실의 형사법적 제조물 책임을 부과할지 아니면 (2) 민법상 제조물책임과 같이 주의의무위반에 대한 입증책임을 부과할지 논란(전지연, 2001; 하태훈, 2002)

- 사람과 인공지능 사이의 책임 귀속 문제: 인공지능의 분석/판단/명령에 따라 행동한 사람이 범죄적 결과로 이어질 경우 누구에게(① 인공지능 ② 설계자/개발자 ③ 사용자 ④ 보험사 등) 형사적 책임을 귀속 시킬지에 대한 논란

**[대응 방안]**

- ➲ 인공지능이 책임 주체로서의 충분한 자격이 있는지에 대한 사회적 합의/공론화 필요
- ➲ 인공지능 행위 결과의 사회적 수용한도에 대한 사회적 합의 필요
- ➲ 인공지능 행위 결과에 대한 이해관계자(설계/개발자, 사용자, 보험사, 제조/서비스 업체 등)의 책임 소재 명확화를 위한 법적 기준 마련

**표 11 인공지능 신뢰 관련 핵심 이슈 및 대응방안**

| 구분 | 이슈 | 대응 방안 |
|---|---|---|
| 산업 | ■ 인공지능이 인간의 일자리를 대체할 수 있다는 우려 | ■ 인공지능으로 새롭게 창출되는 직업군에 대한 연구<br>■ 질 중심의 전문화된 일자리 창출 정책 추진 |
| 기술 | ■ 인공지능 설계시 고려하지 못한 조건의 발현, 기술의 복잡성 증가 등으로 인공지능 시스템의 오류 발생에 대한 우려 | ■ 인공지능 관련 알고리즘의 오류나 오작동의 문제를 최소화할 수 있는 기술개발 및 이를 규제하기 위한 방안 마련<br>■ 인공지능 기술에 의한 오작동을 줄이기 위한 사용자의 법적 책임 소재 명확화 |
| 데이터 | ■ 부정확한 데이터 사용으로 인한 인공지능의 피해 야기 우려<br>■ 개인정보 유출, 프라이버시 침해에 대한 우려 | ■ 해킹 방지 인공지능 서비스 개발<br>■ 데이터 처리 및 저장 방식, 사용 범위 등에 대한 규정 마련 |
| 사회/문화 | ■ 악의적인 의도로 인공지능 활용에 따른 각종 위협에 대한 우려<br>■ 인공지능과의 새로운 관계 형성 의존도 증가로 비인간적, 비사회성에 대한 우려 | ■ 인공지능에의 감정이입을 이용한 인간감정의 조작을 방지하기 위한 투명한 연구개발 윤리 및 가이드라인 마련 필요<br>■ 인공지능 오남용을 사전에 탐지, 대처하는 기술 개발<br>■ 악의적인 인공지능 활용에 대비하여 인공지능 기술 및 시스템 개발 가이드라인과 중대한 위반시 법적 제재 항목의 설정 필요 |
| 윤리 | ■ 인공지능에게 인간 행위의 옳고 그름에 대한 윤리적 판단 기능 부여시 신뢰 문제 | ■ 인공지능에 윤리적 판단을 부여하는 것이 합리적인지에 대한 평가 및 사회적 합의/공론화가 선결 요건<br>■ 인공지능의 자율적 판단관련 알고리즘 개발 단계에서부터 구체적인 가이드라인 마련 필요 |
| 법적책임 | ■ 인공지능의 자율적 판단과 행동에 대한 책임 소재의 불명확성 우려 | ■ 인공지능이 책임 주체로서의 충분한 자격이 있는지에 대한 사회적 합의 필요<br>■ 인공지능 행위 결과의 사회적 수용한도에 대한 사회적 합의 필요 |

## 주요 산업의 인공지능 신뢰 이슈 및 대응 방안

### 자동차 산업

- 인공지능의 적용이 확산됨에 따라 산업현장의 일자리 대체, 자율주행자동차 시스템의 기술적 오류로 인한 사고 발생, 운행중 개인 데이터의 외부 유출우려, 악의적 해킹에 의한 인적/물적 피해 우려 등의 이슈 부각

- 자율주행차의 사고 발생 상황시 탑승자와 보행자중 누구를 보호해야 할지에 대한 윤리적 가치판단의 신뢰문제 야기

- 자율주행차의 도입은 새로운 도로교통법의 제정을 요구하며 교통사고 발생시 탑승자-제조사-설계사/개발자-보험사 간의 새로운 책임 귀속의 설정 문제 야기

- 인공지능 자동차의 윤리적 딜레마(Trolley Problem) 이슈 부각: 위급상황에서 탑승자와 보행자 중 누구를 살릴 것인가에 대한 윤리적 판단의 문제

  * 자율주행자동차가 탑승자와 다수의 보행자중 누구를 구할 것인지에 대한 설문조사 결과 대부분의 응답자들은 희생을 최소화하는 자율주행자동차를 선호하는 공리주의적 가치관을 피력하였으나, 정작 응답자 당사자들은 다수를 살리는 인공지능 알고리즘을 장착한 자율주행차를 타고 싶지 않겠다는 반응을 보임(MIT Technology Review, 2015)

  * 구글의 자율주행차가 사고낸 사례: 2016년 세계에서 자율주행차로 인한 최초의 교통사고 사례로서 당시 현행법상 자율주행차의 운전자를 탑승자로 보기 때문에 미국 법원은 사용자에게 사고책임을 물음

**자율주행자동차의 트롤리 딜레마**


출처: 조선일보(2016)

**[대응 방안]**

① 산업

➲ 자율주행자동차 환경에 맞는 일자리 창출 방안 연구: 예) SW설계, 품질평가/관리, 보안 분야 등

② 기술

➲ 자율주행자동차 SW 결함 탐지 기술 개발 및 오류를 최소화할 수 있는 기술 개발 추진

➲ 자율주행자동차의 오작동으로 인한 사고에 대한 기술 규명 및 법적 책임 소개 규정 마련

③ 데이터

➲ 편향되지 않고 오류가 없는 데이터 사용을 위한 데이터 활용 규정 준수 마련

➲ 자율주행자동차 위치, 운행정보 및 개인 이용자 데이터가 공정하게 다루어질 수 있도록 기술적 솔루션 개발 필요

➲ 자율주행자동차의 해킹 방지를 위한 인공지능 기술 개발

④ 사회/문화

➲ 자율주행자동차 SW가 해커 등 악의적인 이용자의 공격을 방어할 수 있는 보안 솔루션 역량 강화

➲ 해커 등 악의적인 이용자에 대한 민형사적 책임 규정을 인공지능 시대의 상황에 맞도록 개선 및 강화

⑤ 윤리

➲ 자율주행자동차가 윤리적 가치판단의 주체가 될 수 있는지에 대한 선제적 사회적 논의/공론화 필요

➲ 자동차 사고의 잠재적 위험성을 최소화하기 위한 기준에서 사회적 합의 도출

⑥ 법제도

➲ 1단계: 자율주행자동차의 법적 책임 대상자 여부에 대한 우선적인 사회적 논의 필요

➲ 2단계: 사고 발생시 자율주행자동차 알고리즘 설계/개발자, 제조회사, 사용자, 보험회사간 책임 소재에 대한 법적 논의 필요

➲ 3단계: 배상 책임의 범위, 배상 요건 등에 대한 세부 가이드라인 마련

● 의료 산업

- 헬스케어산업에 인공지능의 적용이 본격화되면서 인공지능의 의료진 업무 대체, 시스템적 오류로 인한 잘못된 진단 및 처방, 무검증의 불명확한 의료 데이터 사용에 따른 의료 과실 발생 및 개인 데이터 유출에 의한 사생활 침해 등의 이슈 부각

- 보험사기 등 악의적 이용 위협, 인공지능에 대한 맹목적 신뢰 증가로 인한 의료진에 대한 불신 증가 등의 사회문제 야기 가능

- 인공지능이 자율적 의사결정에 의한 진단 및 처방에 대해 윤리적 가치 판단의 문제와 부득이한 의료사고 발생 시 인공지능, 의사, 설계자/개발자, 보험회사간 책임 귀속 문제 논란 지속 전망

**[대응 방안]**

① 산업
- ➲ 인공지능 헬스 시스템과 의료진의 상호 협력의 업무 시스템 연구

② 기술
- ➲ 인공지능 헬스케어 시스템에 대한 검증 강화 및 시스템 오류 방지 기술 개발
- ➲ 인공지능에 의한 진단/처방을 의료진의 진료 보조 자료로 활용함으로써 최종적인 의사결정을 의료진이 담당

③ 데이터
- ➲ 인공지능이 활용하는 의료데이터의 수집, 관리, 보안 등에 관한 의료DB 활용 가이드라인 마련
- ➲ 인공지능 헬스케어 정보에 대한 해킹 방지 기술 개발

④ 사회/문화
- ➲ 인공지능이 다루는 의료데이터에 대한 보안 시스템 강화
- ➲ 개인 의료데이터 유출, 보험사기 등의 인공지능을 활용한 악의적 이용자에 대한 민법/형법적 책임 부과 기준 마련
- ➲ 인공지능과 의료진간 상생 및 상호협력의 新 의료진료 시스템 구축

⑤ 윤리
- ➲ 헬스케어에서 인공지능 시스템이 윤리적 가치판단을 할 수 있는 주체인지에 대한 선 사회적 합의/논의 필요
- ➲ 인공지능이 윤리적 가치판단의 주체 여부에 대한 인류사회에 미치는 영향

평가 연구 수행

⑥ 법제도

➲ 1단계: 헬스케어 분야 인공지능의 법적 책임 대상자 여부에 대한 우선적인 사회적 논의 필요

➲ 2단계: 의료사고 발생시 인공지능 알고리즘 설계/개발자, 사용자, 병원, 보험회사간 책임 소재에 대한 법적 논의 필요

➲ 3단계: 배상 책임의 범위, 배상 요건 등에 대한 세부 가이드라인 마련

**표 12 인공지능 신뢰 관련 핵심 이슈 및 대응: 자동차와 의료 산업**

| 구분 | | 자동차 | 의료 |
|---|---|---|---|
| 산업 | 이슈 | ■ 인공지능에 의한 고도화된 자동화 시스템으로 일자리 감소 우려 | ■ 인공지능 헬스케어의 확산으로 의료진의 업무 대체 우려 |
| | 대응 | ⇨ 자율주행차 산업환경에 적합한 일자리 창출 연구 | ⇨ 인공지능과 의료진의 상호 협력의 업무 시스템 모색 |
| 기술 | 이슈 | ■ 자율주행자동차 시스템의 기술적 오류로 인한 사고 발생 가능성 우려 | ■ 인공지능 헬스케어 시스템의 오류로 인한 잘못된 진단 및 처방에 대한 우려 |
| | 대응 | ⇨ 자율주행자동차 SW의 결함 탐지 및 오류 최소화 기술 개발<br>⇨ 자율주행자동차의 오작동에 의한 사고결과의 기술규명 및 법적 책임소재 규정 | ⇨ 인공지능 헬스케어 시스템 오류 방지 프로그램 개발<br>⇨ 인공지능의 진단/처방을 의료진이 진료보조 자료로 활용하여 환자에게 최종 진단과 처방 수행 |
| 데이터 | 이슈 | ■ 오류가 있는 부정확한 자동차 운행 DB 사용에 따른 인공지능 분석의 잘못된 결과 생성 가능성 우려<br>■ 자율주행자동차 운행시 실시간 차량정보 데이터의 외부 유출로 인한 프라이버시 침해 우려<br>■ 자율주행자동차의 공유차량 증가로 인한 이용자 개인 데이터 관리 부실로 인한 프라이서비스 침해 우려 | ■ 검증되지 않은 의료 데이터 사용으로 의료 과실 발생 우려<br>■ 개인 의료 데이터 유출, 부정한 열람 등으로 인한 사생활 침해 우려 |
| | 대응 | ⇨ 오류 없는 정확한 데이터 사용 규정 준수 마련<br>⇨ 자율주행자동차 해킹 방지 인공지능 기술 개발 | ⇨ 인공지능 환경에 맞는 의료데이터 생성, 관리, 보안 등에 관한 의료DB 가이드라인(안) 마련<br>⇨ 인공지능 헬스케어 해킹 방지 기술 개발 |
| 사회/문화 | 이슈 | ■ 악의적인 해킹 공격으로 인한 자동차 테러 및 사고 발생 위협 | ■ 개인 의료데이터를 보험사기 등 악의적 이용에 대한 위협<br>■ 인공지능 헬스케어 의존도/신뢰 증가로 의료진에 대한 불신 증가 |

| | | | |
|---|---|---|---|
| | | | 우려 |
| | 대응 | ⇨ 자율주행차 SW의 보안 시스템 강화 기술 개발<br>⇨ 해커 등 악의적인 이용자에 대한 형법/민법적 책임 부과 기준 마련 | ⇨ 인공지능 헬스케어 데이터 보안 시스템 강화 기술 개발<br>⇨ 악의적 이용자에 대한 형법/민법적 책임 부과 기준 마련<br>⇨ 인공지능과 의료진간 상호협력의 新 의료진료 시스템 구축 |
| 윤리 | 이슈 | ■ 자율주행자동차의 자율적 의사결정에 대한 윤리적 가치판단의 신뢰 여부<br>- 긴박한 사고 발생 위험시 누구(탑승자 or 보행자)를 보호해야 할 것인지? | ■ 인공지능이 의료 진단 및 처방에 대한 가치판단의 윤리적 문제 |
| | 대응 | ⇨ 자율주행자동차의 합리적 판단기준 부여에 대한 先 사회적 합의/공론화 필요 | ⇨ 인공지능 헬스케어 시스템이 윤리적 가치판단의 주체여부 및 가차판단 부여에 대한 先 사회적 합의/공론화 필요 |
| 법제도 | 이슈 | ■ 자율주행자동차의 사고발생시 배상책임 귀속 문제<br>■ 자율주행차에 대한 제조물 책임 부과 문제 | ■ 인공지능이 의료사고 발생시 책임귀속 문제 |
| | 대응 | ⇨ 자율주행자동차에 대한 先 법적책임 주체 여부 판단 기준 마련<br>⇨ 자율주행자동차 알고리즘 설계/개발자, 사용자, 제조업체, 보험사간 배상책임 소재의 법규정 마련<br>⇨ 배상 책임의 범위 및 구체적인 배상 요건에 대한 가이드라인 마련 | ⇨ 인공지능 헬스케어 시스템의 설계/개발자, 의료진, 병원, 보험사간 배상책임 소재의 법규정 마련<br>⇨ 배상 책임의 범위 및 구체적인 배상 요건에 대한 가이드라인 마련 |

**소비자 프라이버시 보호 원칙(Consumer Privacy Protection Principles) 발표(2014년)**

◘ 목적: 인공지능 활용에 따른 차량정보 및 개인 데이터의 외부 유출 문제 이슈의 선제적 대응을 위해 자동차 제조업체 합의(Alliance of Automobile Manufactures)로 개인정보보호 원칙 제시

◘ 7가지 프라이버시 보호 원칙

① **투명성(Transparency)**: 차량에서 생성, 기록, 저장되는 정보를 제조사가 사용할 경우 이를 소비자에게 투명하게 공개

② **선택권(Choice)**: 차량에서 생성, 기록, 저장되는 정보의 공개 여부(제조사의 사용여부)는 소비자에게 있음

③ **맥락의 일관성(Respect for Context)**: 차량에서 생성, 기록, 저장되는 정보는 수집 목적

에 맞도록 일관성있게 처리

④ **수집정보는 최소화, 비식별화 및 보존(Daga Minimization, De-Identification & Retention)**: 차량에서 생성, 기록, 저장되는 정보는 최소한으로 수집돼야 하며, 비식별화 조치를 거친 후 보존

⑤ **데이터 보안(Data Security)**: 차량에서 생성, 기록, 저장되는 정보는 무단 접근과 절취 등이 발생하지 않도록 보호

⑥ **무결성과 접근성(Integrity & Access)**: 차량에서 생성, 기록, 저장되는 정보는 무결성이 유지돼야 하며 합리적인 보호조치를 하여 이를 소비자에게 제공

⑦ **책임의무(Accountability)**: 차량에서 생성, 기록, 저장되는 정보를 활용하는 자동차제조업체와 이해당사자들은 소비자 프라이버시 보호 원칙을 준수할 것이라는 보증 방안 마련

자료: https://autoalliance.org/wp-content/uploads/2017/01/Consumer_Privacy_Principlesfor_VehicleTechnologies_Services.pdf, NIA(2016)

- 산업, 기술, 데이터, 사회/문화, 윤리, 법제도 측면에서 인공지능 신뢰 이슈 및 대응 방안 제시
- 인공지능 기술의 도입에 따른 일자리 대체 우려, 인공지능 시스템 오류로 인한 피해 발생 우려, 개인정보 유출 및 프라이버시 침해, 악의적 인공지능 활용에 의한 각종 위협, 인공지능에 윤리적 가치판단 부여 및 법적 책임 귀속 문제 등이 신뢰관련 핵심 이슈로 부각
- 인공지능에 대한 걱정/우려/불신 등을 해소하기 위해서는 인공지능 설계단계에서 오남용 방지 및 개발 가이드라인 기준 마련이 필요하며, 프라이버시 침해 등에 대응한 기술개발 필요
- 또한 인공지능의 자율적 의사결정에 의한 윤리적 판단 부여 관련 사회적 합의가 우선적으로 필요하며, 사고 발생시 누구에게 책임을 귀속시킬 것인가에 대해 인공지능의 인적/물적 대상에 대한 법리적 검토 및 책임을 부여하기 위한 조건에 대한 사전 연구 필요

# ※ 참고문헌


강수택, 사회적 신뢰에 관한 이론적 시각들과 한국 사회, 사회와 이론, 2003.

과기정통부, 지능정보사회 중장기 종합대책, 2017.

관계부처합동, 4차 산업혁명 대응 계획, 2017.

김대원, 인공지능이 쓴 기사에 대한 소비 선택의도에 영향을 미치는 요인, 한국방송학보, 제31권 제6호, 2017.

양희태, 인공지능의 위험성에 대한 우려로 제정된 아실로마 인공지능 원칙, 2017.

윤지영 외, 법과학을 적용한 형사사법의 선진화 방안, 연구총서 15-B-16, 한국형사정책연구원, 2015.

임석순, 형법상 인공지능의 책임귀속, 형사정책연구, 제27권 제4호, 2016.

전지연, 형벌론적 관점에서 본 형법적 제조물책임의 필요성, 형사정책, 제13권 제1호, 형사정책학회, 2001.

조설일보, 충돌 때 탑승자, 보행자, 누굴 살리나?...無人車 딜레마, 2016.

중앙일보, 점점 커지는 일자리 감소 우려, 2017. http://news.joins.com/article/18457148

Colman, J. S., Foundation of Social Theory, 1990, Cambridge: The Belkap Press of Harvard University Press.

Gambetta, D. (ed.), Trust: Making and Breaking Cooperative Relations, 1988, N Y.: Basil Blackwell.

IBM, Learning to trust artificial intelligence systems: Accountability, compliance and ethics in the age of smart and machines, 2016.

Luhmann, N., Familiarity, Conficence, Trust, Trust, D. Gambetta (ed.), 1988, N. Y.: Basil Balckwell.

Misztal, B. A., Trust in Modrn Societies, 1996, Cambridge: Pilicy Press.

MIT Technology Review, Why Self-Driving Cars Must be Programmed to Kill, 2015

NIA, 중국의 인공지능 전략: 차세대 인공지능 발전계획을 중심으로, 2017.

NIA, AI 연구개발과 활용 촉진을 위한 'AI 개발 가이드라인(안)', 2017.

NIA, 프랑스의 인공지능(AI) 전략: 인공지능의 사회적, 경제적 영향 전망을 중심으로, 2017.

NIA, 美 연방 자율주행차 가이드라인: 주요내용 및 시사점, Special Report 2016-3, 2016.

SPRI, 지능정보사회 대응을 위한 법제도 조사연구, 2017.

The Science Times, 인공지능 오작동, 새로운 위험, 2017.

https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/federal_automated_vehicles_policy.pdf

https://futureoflife.org/ai-principles

https://autoalliance.org/wp-content/uploads/2017/01/Consumer_Privacy_Principlesfor_VehicleTechnologies_Services.pdf

## 저자소개

김 문 구 ETRI 미래전략연구소 기술경제연구본부 기술정책연구그룹 책임연구원
e-mail: mkkim@etri.re.kr Tel. 042-860-1182

박 종 현 ETRI 미래전략연구소 기술경제연구본부 산업전략연구그룹 선임연구원
e-mail: stephanos@etri.re.kr Tel. 042-860-1081

**인공지능과 신뢰(Trust): 이슈 및 대응 방안**

발 행 인 : 한 성 수

발 행 처 : 한국전자통신연구원 미래전략연구소 기술경제연구본부

발 행 일 : 2017년 12월

ETRI 한국전자통신연구원
미 래 전 략 연 구 소

305-700 대전광역시 유성구 가정로 218
전화 : (042) 860-1182, 팩스 : (042) 860-6504

ETRI 미래전략연구소 www.etri.re.kr