

# 인간과 인공지능의 공존(共存)을 위한 당면 과제

2019. 05. 21.  
제43회 SPRi Forum

## 추 형 석

[hchu@spri.kr](mailto:hchu@spri.kr)

소프트웨어정책연구소 선임연구원  
기술·공학 연구실

소프트웨어 중심사회의 Think Tank  Software Policy & Research Institute



# 목 차

1. 토론회 배경

2. 주요 토론 내용

# 1. 토론회 배경

# Will AI enhance or hack humanity? (1/2)

- 미래 인공지능과 인류에 대해 논하는 좌담회
- 역사학자이자 여러 베스트셀러를 집필한 유발 하라리, 심층학습의 석학인 스탠포드대학 교수 페이페이 리가 참석



# Will AI enhance or hack humanity? (2/2)



유발 하라리 (Yuval Harari)

- 인공지능의 발전이 인간에게 악영향을 미칠 것이라 강조

$$B \times C \times D = H$$

Biology, computer, and data  
are able to hack humanity.

- 나보다 더 나를 잘 이해하는 알고리즘이 수 많은 선택에 관여



페이페이 리 (Fei-Fei Li)

- 인공지능이 오픈사이언스의 특성으로 인해 많은 사람들이 쉽게 접근
- 인공지능이 인간을 해킹할 수 있는 능력은 아직 먼 미래의 이야기
- 그러나 그 대비를 위해 인간 중심의 인공지능 활용을 위한 다학제적인 협업이 필요

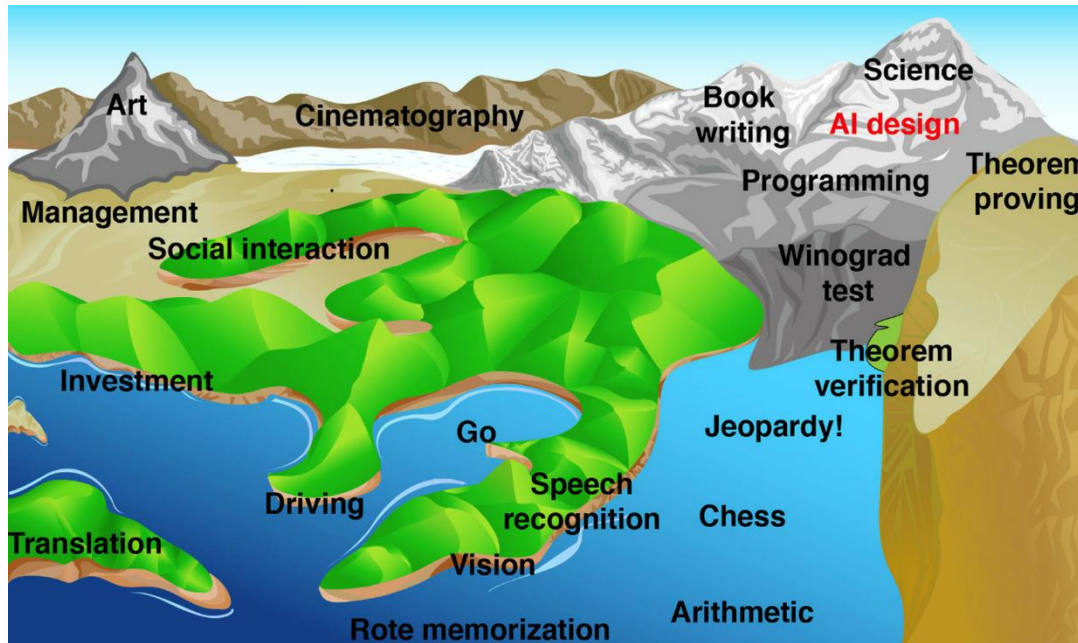
# 인공지능의 명과 암

- **현대 인공지능은 심층학습(Deep Learning)으로 세 번째 황금기를 맞이함**
  - 데이터를 학습하여 패턴을 인식하고 미래를 예측
  - (의료) 폐암, 유방암, 당뇨망막변증 등 질병 예측, 가상 간호 조무, 정밀 의료
  - (교통) 자율주행차, 지능형 교통 시스템, 커넥티드 카
  - 스마트 스피커, 지능형 비서, 맞춤형 광고 및 제품 추천 등
- **인공지능의 고도화로 인한 잠재적 위험 부상**
  - (사회) 일자리 문제, 부의 양극화 심화, 인공지능 윤리
  - (기술) 인공지능 시스템의 신뢰성과 공정성, 데이터와 알고리즘의 편향
  - 인공지능 기반 무기체계, 악의적인 테러, 고도화된 인공지능이 인간을 지배
- **“인공지능을 인류에 이롭게 활용하자”**
  - 국제사회는 인공지능을 인류에 이롭게 활용하자는 공감대 형성
  - 23대 아실로마 AI 원칙을 공개하여 그 방향을 제시
  - 인문사회과학 분야와 연구자 및 기술자의 긴밀한 협업을 강조

# 범용 인공지능과 초지능

## ● 범용 인공지능의 부상

- 범용 인공지능 혹은 사람 수준의 인공지능은 좁은 인공지능의 반대되는 개념
- 현재 인공지능 기초원천 연구는 현대 심층학습을 한계를 극복하여 더 넓은 범용성을 확보하는 방향으로 추진 (메타 학습, 전이 학습 등)
- 범용 인공지능의 개발은 재귀적으로 학습하여 진화하는 초지능의 특이점으로 해석



# 인간과 인공지능의 공존(共存)

---

- 시점의 분리 : 현재와 미래
- (현재) 우리가 인공지능을 활용함에 있어 당면하고 있는 과제를 해결하기 위한 방안
- (미래) 범용 인공지능이나 초지능의 출현에 대비하기 위한 논의



# (현재) 인공지능 산업 활성화

- 인공지능 기술을 타 산업에 적용하여 자동화하거나 의사결정을 지원하는 산업적 활용의 당면 과제
  - 인공지능의 역효과에 대한 공론화와 이에 대한 대책이 필요
  - 신산업과 기존 사업과의 갈등 해소
  - 개인정보보호법 등 각종 법제도와 데이터의 활용
  - 인공지능 알고리즘의 신뢰성과 공정성
- (예시) 자율주행차의 도입
  - 사고 발생 시 책임 소지, 보험료 산정 기준
  - 트롤리 문제 등 윤리적 판단에 대한 기준
  - 기존 운수 산업과의 갈등 해소
  - 자율주행차가 인간보다 합리적이 다는 사실을 다양하게 증명

# (미래) 범용 인공지능 시대에 대한 대비

---

- **국제사회의 움직임 : 범용 인공지능에 대한 대비**
  - 인간 중심의, 인간에 이로운 인공지능을 강조
  - 미래 삶 연구소(FLI) 설립자 중 한명인 MIT 맥스 테그마크는 TED 강연에서 인공지능을 이롭게 활용하기 위해 방향성을 강조
  - 인공지능의 악의적인 활용에 대한 대비, 인공지능을 둘러싼 주체들 간의 의무 제시 등 다양한 제안이 나오는 중
  - 아실로마 AI 원칙 23가지 제시 (미래 삶 연구소, 2017)
- **미래 인공지능에 대한 우리의 대응**
  - 국제사회의 움직임에 동조할 수 있는 대응 필요
  - 인공지능의 어두운 부분, 한계에 대한 공감대 형성 필요

## 2. 주요 토론 내용

# 철학 (서울대학교 홍성욱 교수)

---

Q1. 인공지능 알고리즘의 편향에 대비하기 위한 “알고리즘 시민권”의 간략한 내용과 그 필요성은?

Q2. 아실로마 AI 원칙에 대한 비평과 당위성, 그리고 우리 사회가 지향해야 하는 바는?

# 윤 리 (서울여자대학교 김명주 교수)

---

**Q1. 인공지능 연구자들 혹은 미래의 인공지능 연구자들이 인공지능의 윤리적 개발에 대해 명확히 인식하고 실천하기 위해 가장 시급한 것은 무엇입니까?**

**Q2. 작년에 발표한 지능정보사회 윤리 헌장에 대한 개괄적인 소개와 향후 방향에 대해서 설명해 주시기 바랍니다.**

# 법·제도 (고려대학교 김기창 교수)

---

Q1. 자율주행차와 로봇에 대한 법적 권리와 의무를 설정하는 방안에 대해서 간략히 설명해 주시기 바랍니다. 또한 법·제도적 관점에서 자율주행차가 해결해야 하는 가장 시급한 현안은?

Q2. “AI 슈퍼파워”의 저자 카이푸리는 급성장한 중국의 인공지능 생태계를 검투장에 비유하며 기업가 정신을 강조했습니다. 우리나라 역시 인공지능을 활용한 창업 생태계가 강화되고 있는 상황인데, 더 많은 도전과 일자리를 창출하기 위해 시급히 해결해야하는 문제는?

# 거버넌스 (카이스트 전길남 명예교수)

---

**Q1. 과거 인터넷 거버넌스와 인공지능 거버넌스의 차이는 무엇인지?**

**Q2. 현재 우리는 인공지능 연구개발과 산업적 활용, 인재 양성 등에 초점을 맞추고 있습니다. 인공지능과 관련된 사회적 합의를 모색하기 위한 거버넌스가 중요한 화두로 부상했는데, 우리가 어떠한 부분에 방점을 두어야하는지 논의 바랍니다.**

# 신 학 (강남대학교 김흡영 명예교수)

---

Q1. 신학의 관점에서 현대 인공지능을 바라보는 해석은?

Q2. 인공지능 시대에 종교의 역할은 무엇인가?



# 기술 (인공지능연구원 김진형 원장)

---

**Q1. 인공지능의 기술패권 경쟁은 미국과 중국의 양강 구도가 극명해진 상황에서 우리의 돌파구는 무엇인가?**

**Q2. 현대 인공지능의 발전상을 볼 때, 인공지능의 겨울이 다시 올 것이라는 예측에 대해서 입장은?**

# (미래) 범용 인공지능 시대에 대한 대비

---

- Q1. 어떠한 형태로든 초지능이 개발될 경우 인간의 존재론적 가치는 어떻게 변화할 것인가?
- Q2. 초지능의 출현에 대비해 우리사회가 공론화하고 지혜를 모아야 하는 가장 시급한 문제는 무엇인가?
- Q3. 범용 인공지능, 초지능에 대한 대중의 인식 개선과 적극적인 참여를 도모할 수 있는 방향은 무엇인가?
- Q4. 인공지능의 신뢰성을 확보하기 위한 주요국의 동향과 접근 방법은?
- Q5. 미래 인공지능의 양면성에 대해 논의

# 감사합니다.

소프트웨어 중심사회의 Think Tank  **SPRI** Software Policy & Research Institute