Korea Centre for the Fourth Industrial Revolution

한국 (경기도)

4차산업혁명센터

경기도

Master's Insight

"AI, 인류의 구원인가 파멸인가?
(Will AI Save the World or End it?)

'AI의 대부' 제프리 힌튼(토론토대 교수)과의 인터뷰

2025년 7월

C4IR Korea Global Intelligence Hub (koreago.net)

본 자료는 2024년 노벨물리학상 수상자이자 'AI의 대부'로 불리는 제프리 힌튼(Jeoffrey Hinton) 캐나다 토론토대 컴퓨터과학과 명예교수가 캐나다 TVO 방송과 2025년 4월 인터뷰한 내용을 한국(경기도) 4차산업혁명센터가 번역하여 요약 정리한 것입니다(자료원 : TVO Today 유튜브 채널). 대담자의 견해는 전적으로 대담자 개인의 의견이며, 한국(경기도) 4차산업혁명센터의 공식 입장이 아닙니다.

Geoffrey Hinton: Will AI Save the World or End it? | The Agenda - 원본 동영상 보기

- **AI는 인간의 뇌를 모방한 시스템으로, 엄청난 가능성을 지닌 동시에 인류 멸종까지 초래할 수 있는 실존적 위협이기도 하다.**

- **AI 안전성을 확보하려면 지금 당장 글로벌 공감대를 형성하고, 대기업과 정부가 공동으로 규제와 연구에 나서야 한다.**

- **AI는 의료와 교육 등 삶의 질을 획기적으로 높일 수 있는 도구인 만큼, 위험을 통제하며 책임 있게 발전시켜야 한다.**

요약 (Executive Summary)

주제: 인공지능의 약속과 위협, 그리고 사회적 대응 방안
출처: TVO 방송 인터뷰 (진행자: Steve Paikin)
출연자: Geoffrey Hinton (AI의 대부, 토론토대학교 명예교수, 노벨 물리학상 수상자)

---

1. 🧠 인공지능의 본질과 연구 동기

신경망의 원리:
인간의 뇌는 뉴런(신경세포) 간 연결 강도를 변화시키며 학습함.
AI도 이 원리를 모방한 인공신경망으로 구성되어, 데이터 기반으로 스스로 학습함.

연구를 선택한 이유:
인간을 이해하려면 뇌를 이해해야 하며, **"뇌는 아직 미지의 영역"**이라는 인식이 연구의 출발점.

---

2. ⚠ AI의 위험: 단기와 장기

단기 위험 (이미 현실화):
AI를 악의적으로 활용하는 사용자들:
- 예) 선거 조작, 딥페이크, 조작 영상, 개인 정보 타겟팅

피싱 공격 1,200% 증가 (2023~2024):
- LLM 기반의 정교한 언어 표현력 탓에 구분이 어려워짐

장기 위험 (실존적 위협):
AI가 인간보다 더 똑똑해질 가능성은 대부분의 전문가들이 동의

- 시점은 1년~20년까지 다양한 의견

인류 멸종 가능성도 존재 (확률: 10~20%)
- 바이러스 설계 등 다양한 시나리오가 가능
- “우리는 그 이후 어떤 일이 벌어질지 알지 못한다”는 점이 가장 큰 문제

---

3. 🏛 사회적·정치적 대응의 필요성

우선순위: 사회적 합의(consensus) 형성
기후변화 대응과 유사:
“문제가 실제로 존재한다는 데 대한 대중의 공감”이 전제되어야 정책 실행 가능

대기업의 책임:
AI 기업들은 **단기 이익보다 안전성 연구에 최소 1/3 이상 자원 투자** 필요
현재는 턱없이 부족한 수준

국제 협력 가능성:
AI가 인류 전체를 위협하는 만큼, 국가 간 협력은 충분히 가능
- “중국도, 트럼프도 멸종을 원하진 않는다”

---

4. AI의 긍정적 가능성

의료 분야:
AI 주치의는 **수억 명의 데이터를 기반으로 정밀한 진단** 가능
의사와 AI가 협업하면 **진단 오류율 현저히 감소**

교육 분야:
AI 튜터는 학습자의 **오해나 약점을 정확히 파악해 개인화된 학습 제공**
인간 과외보다 **3~4배 더 효율적인 교육 기대**
향후 10년 이내 현실화 가능성 높음

---

5. 📓 개인사와 철학적 시사점

노벨상과 기부:
노벨 물리학상 수령액의 절반을 **‘Water First’에 기부**
- 캐나다 원주민 지역의 안전한 식수 공급 프로젝트

■ “깨끗한 물은 인간의 기본권”

일론 머스크와의 충돌:
트위터/X 상에서 갈등
■ 머스크의 “과학 기관 파괴”와 “정부 인력 대량 해고”를 **“끔찍한 일”**이라 지적

Musk 측 반응: “무의미한 비판”
Hinton: “나는 대화 대신 행동과 연구에 집중하겠다”

📌 결론

Geoffrey Hinton 교수는 **AI의 놀라운 잠재력**을 인정하면서도,
그 **실존적 위협에 대한 국제적 경계와 윤리적 통제의 필요성**을 강하게 호소한다.
AI의 발전은 더 이상 기술자의 문제가 아니라,
인류 전체가 함께 논의하고 책임져야 할 정치적·사회적 과제라는 것이 그의 핵심 메시지다.

인터뷰 전체 번역본 (주요 내용 위주)

(“AI의 대부” Geoffrey Hinton 교수, TVO 출연)

STEVE:
오늘의 게스트는 "AI의 대부(Godfather of AI)"로 잘 알려진 분입니다.
최근에는 **인공지능 분야의 선구적 연구**로 **노벨 물리학상**을 수상하기도 했죠.
또한 **일론 머스크의 심기를 거스른 인물**이기도 합니다.
그 이야기 역시 놓치지 않고 여쭤보겠습니다.
이분은 토론토대학교 컴퓨터과학과의 명예교수인 **Geoffrey Hinton 교수님**입니다.
오늘 이 자리에 모셔서 **AI의 약속과 위험성**에 대해 깊이 있는 대화를 나눠보겠습니다.
스튜디오에 다시 오셔서 정말 반갑습니다.
GEOFFREY:
초대해 주셔서 감사합니다, Steve.

🎥 스톡홀름 노벨상 수상 장면

STEVE:
먼저, 당신 인생에서 가장 멋졌던 날 중 하나의 장면을 짧게 보시죠. 셸던, 영상 부탁드립니다.

(클립 재생) 🎵
지금 보신 장면은 **스웨덴 국왕 칼 구스타프 16세**께서
스톡홀름에서 노벨상을 수여하시는 모습이고, 바로 당신이 수상자였습니다.
그 감동은 언제쯤 사라지던가요?
GEOFFREY:
글쎄요, 아직도 완전히 가시지는 않았습니다.
STEVE:
그만큼 특별했던 날이었군요.
GEOFFREY:
그럼요. 특히 저는 **물리학을 전공하지 않았는데도** 물리학상을 받았다는 점에서 더욱 놀라웠죠.
STEVE:
어떻게 그런 일이 벌어진 건가요?
GEOFFREY:
제 생각에 위원회는 **AI 분야의 과학적 발전에 노벨상을 수여하고 싶었던 것 같습니다.**
요즘 과학계의 가장 뜨거운 분야가 바로 AI이니까요.
그래서 아마 **물리학상의 의미를 조금 확장한 것**이라 생각합니다.
STEVE:
그들에게 "저 물리학 안 해요"라고 말했나요?
GEOFFREY:
말하긴 했죠. 그런데 뭐, 금메달이니까요. (웃음)
STEVE:
그 메달은 어디에 보관 중인가요?
GEOFFREY:
그건 비밀입니다. 누가 훔쳐갈 수도 있으니까요.
금 6온스로 약 15,000달러쯤 하거든요.

🧠 "신경망"은 어떻게 작동하나?

STEVE:
수상 이유는 이렇게 설명되어 있습니다:
"인공 신경망을 통한 기계 학습의 기초적 발견과 발명."
그런데 대중들은 이게 무슨 말인지 잘 모릅니다.
이걸 쉽게 설명해 주시겠어요?
GEOFFREY:
좋습니다.
우리 뇌에는 **뉴런**이라는 뇌세포들이 수없이 존재하고, 이들은 서로 연결되어 있습니다.
새로운 것을 배울 때 일어나는 일은 **바로 이 연결의 강도가 조정되는 것**입니다.
즉, 뇌는 어떤 연결을 강하게 하고 어떤 연결은 약화시켜야 할지를 스스로 판단하죠.
그리고 그 판단 방식—어떻게 연결 강도를 바꿀지 결정하는 룰이 무엇인지를 아는 것이 핵심

입니다.
그 메커니즘을 컴퓨터로 모방해서 **인공 신경망**을 구성할 수 있습니다.
그리고 우리는 이제 압니다.
그렇게 하면, 매우 똑똑한 시스템이 됩니다.

🔍 왜 이 연구를 선택했나?

STEVE:
수많은 과학 분야 중 왜 이걸 선택하셨나요?
GEOFFREY:
가장 흥미로운 분야였기 때문이죠.
인간을 이해하려면 뇌를 이해해야 합니다.
그리고 아직도 우리는 뇌가 어떻게 작동하는지 충분히 모릅니다.
그래서 더더욱 중요하다고 생각합니다.

⚠ AI의 위험: 단기와 장기

STEVE:
최근 당신은 AI의 위험성에 대해 경고해 왔습니다.
구글에서도 그런 우려로 퇴사하셨다고 들었어요.
위험을 **단기와 장기**로 나눠 설명해 주시겠어요?
GEOFFREY:
네, 두 가지로 나눌 수 있습니다.

단기 위험:
악의적인 사용자가 AI를 오용하는 것입니다.
예를 들어, 사람들의 데이터를 수집해 조작된 영상으로 선거에 영향을 준다든지,
사이버 공격을 정교화한다든지 하는 일이 이미 벌어지고 있습니다.
📈 참고로 2023년~2024년 사이, 피싱 공격은 **1200% 증가**했습니다.
AI가 문법적으로 완벽한 피싱 이메일을 만들 수 있게 되었기 때문입니다.

장기 위험:
AI가 인간보다 **더 똑똑해지는 순간**,
우리 인간을 필요 없는 존재로 여길 가능성입니다.
더 지능적인 존재가 덜 지능적인 존재에게 지배받는 사례가 얼마나 있을까요?
거의 없죠. 그게 바로 두려운 점입니다.

STEVE:

그 가능성은 얼마나 되나요?
GEOFFREY:
정확히 알 수는 없지만, **10~20%는 된다고 봅니다.**
단순히 영화 '터미네이터' 같은 상상은 아닙니다.
예를 들어 **치명적인 바이러스를 설계해 인류를 제거할 수도 있습니다.**

🏛 정부와 기업은 무엇을 해야 하나?

STEVE:
그렇다면 우리는 무엇을 해야 할까요?
GEOFFREY:
우선, **이 문제가 진짜라는 데 대한 사회적 합의(Consensus)를 형성해야 합니다.**
기후변화 대응과 비슷하죠.
그다음에야 정부가 행동할 수 있습니다.
그리고 **대형 AI 기업들이 수익만 쫓지 않고,**
전체 자원의 최소 1/3은 안전 연구에 투자해야 합니다.
하지만 현재는 연구가 턱없이 부족합니다.

🌍 국제 협력의 가능성과 한계

STEVE:
국제 협력은 가능할까요?
GEOFFREY:
가능하다고 봅니다.
왜냐하면 어떤 국가도 **자신이 파괴되길 바라지는 않으니까요.**
중국도, 트럼프도 마찬가지입니다.
그래서 AI가 인류를 위협할 가능성에 대해서는 **국경을 넘어 협력**이 가능하다고 생각합니다.

🧠 일론 머스크와의 갈등

STEVE:
이제 일론 머스크 얘기로 넘어가 보죠.
그와 사이가 별로 좋지 않다는 얘기가 있더군요.
GEOFFREY:
그가 AI의 위협을 심각하게 본다는 점에서는 저도 동의합니다.
그가 만든 전기차나 스타링크 같은 기술도 훌륭하다고 생각합니다.
하지만 그가 정부 공무원들을 **무작위로 해고하고**,

그들을 **게으르고 부패했다고 몰아붙이는 행태**는 **끔찍하다고** 생각합니다.
제가 그를 **왕립학회에서 퇴출시키자**고 트윗했더니,
그는 저를 "비겁하고, 상을 탐내는 어리석은 자"라고 반응했죠.
그래서 더 이상 대화는 이어가지 않았고, 제 친구 Yann LeCun이 대신 반박했습니다.

노벨 상금과 '물 프로젝트'

STEVE:
노벨상 상금의 절반, 약 35만 캐나다 달러를 "Water First"에 기부하셨다고요?
GEOFFREY:
맞습니다.
"Water First"는 **캐나다 원주민 지역에 안전한 식수를 공급하는 기술 교육 단체**입니다.
제가 페루에 입양된 아이를 만나러 갔을 때,
그곳의 **수돗물이 치명적일 정도로 오염되어 있다는 걸 경험**했어요.
그래서 안전한 물이 없는 삶이 얼마나 비극적인지 체감했습니다.
캐나다 같은 부국에서조차
20%의 원주민 공동체가 깨끗한 물을 공급받지 못하는 현실은
정말 부끄러운 일이라고 생각합니다.

앞으로의 계획

STEVE:
앞으로의 계획은요?
GEOFFREY:
은퇴하려 했지만 잘 안 되고 있네요. (웃음)
제가 77세인데 아직도 일이 많습니다.
STEVE:
당신에겐 아직 몇 개의 챕터가 남아있는 것 같네요.
GEOFFREY:
Steve, 당신도 77세치고 멋지게 보이시네요. (웃음)

낙관적 메시지

STEVE:
이 대화의 마지막, **조금은 희망적인 메시지**를 주실 수 있을까요?
GEOFFREY:
물론입니다.

AI는 **의료와 교육 분야에서 엄청난 긍정적 가능성**을 가지고 있습니다.
예를 들어, AI 기반 주치의는 **1억 명의 환자 데이터를 기억하고,**
가족력까지 고려한 정확한 진단을 할 수 있습니다.
교육에서는, **AI 튜터가 수백만 명의 학습 데이터를 기반으로,**
학생 개개인의 오해를 정확히 짚어주는 맞춤형 교육을 제공할 수 있죠.
아마 향후 10년 이내에 현실화될 겁니다.

STEVE:
Geoffrey Hinton 교수님, 오늘 정말 귀한 말씀 감사합니다.
언젠가 다시 일론 머스크와 손을 잡고 문제 해결에 나서실지도 모르겠네요.
GEOFFREY:
그럴 가능성은 낮다고 봅니다. (웃음)
STEVE:
지금까지 **토론토대학교 컴퓨터과학과 명예교수이자 노벨상 수상자 Geoffrey Hinton** 교수였습니다. 감사합니다.