



인공 일반 지능(AGI)으로의 전환 거버넌스: 유엔 총회 긴급 고려 사항

유엔 총회 의장단(UNCPGA) 보고서

집행 요약

AI 시스템은 다양한 인지 작업에서 인간 지능과 동등하거나 능가하는 시스템으로 특징지어지는 인공 일반 지능(AGI)을 향해 빠르게 발전하고 있습니다. 역사상 가장 큰 규모의 재정 투자가 전례 없는 연구 개발 노력을 주도함에 따라, 업계 리더와 전문가들은 AGI가 금세기 안에 출현하여 인류에게 엄청난 혜택을 가져다줄 수 있다고 예측합니다. 이러한 혜택 중 AGI는 공중 보건 관련 과학적 발견을 가속화하고, 많은 산업을 혁신하고 생산성을 향상시키며, 지속 가능한 개발 목표 달성을 기여할 수 있습니다.

그럼에도 불구하고 AGI는 독특하고 잠재적으로 파국적인 위험을 초래할 수도 있습니다. 기존 AI와 달리 AGI는 인간의 감독 없이 자율적으로 유해한 행동을 실행하여 돌이킬 수 없는 영향, 첨단 무기 시스템의 위협, 중요 인프라의 취약성을 야기할 수 있습니다. AGI의 엄청난 혜택을 누리려면 이러한 위험을 반드시 완화해야 합니다.

이러한 글로벌 과제를 효과적으로 해결하기 위해서는 유엔의 지원을 받는 즉각적이고 조율된 국제적 조치가 필수적입니다. 이러한 조치는 AGI의 이점과 위험, 그리고 잠재적인 글로벌 AGI 관측소, 안전하고 신뢰할 수 있는 AGI 인증 시스템, 유엔 AGI 협약, 국제 AGI 기구 설립을 논의하기 위한 특별 유엔 총회에서 시작되어야 합니다. 선제적인 글로벌 관리가 없다면 국가 및 기업 간의 경쟁은 위험한 AGI 개발을 가속화하고, 보안 프로토콜을 약화시키며, 지정학적 긴장을 악화시킬 것입니다. 조율된 국제적 조치는 이러한 결과를 방지하고, 안전한 AGI 개발 및 사용, 혜택의 공평한 분배, 그리고 글로벌 안정을 촉진할 수 있습니다.

I. 서론

AI의 발전 속도는 최근 몇 년, 특히 최근 몇 달 동안 매우 빨랐으며, 더욱 가속화될 수 있습니다. 이는 부분적으로 AI 기업들이 더욱 능동적이고 자율적인 AI 에이전트를 만드는 데 막대한 투자를

하고 있고, 가장 강력한 AI 모델이 AI 연구 자체를 발전시키는 데 점점 더 많이 사용되고 있기 때문입니다. 이러한 AI 능력의 향상은 대부분의 인지 작업에서 인간의 성능과 동등하거나 능가하는 AI 시스템인 "인공 일반 지능"(AGI)으로 이어질 것이라는 것이 널리 예상됩니다.

AGI 출현 시기에 대한 의견 불일치가 있지만, 이 패널의 모든 전문가들은 AGI가 금세기 안에 개발될 가능성이 높다고 믿습니다. AI 기업들은 AGI를 매우 빠른 시일 내에 달성하기 위해 수백억 달러를 투자하고 있으며, 이는 인류 역사상 가장 큰 연구 개발 노력입니다. 민간 부문은 훨씬 더 안전한 기술을 개발할 책임이 있으며, 그렇게 할 인센티브를 가져야 합니다. 그러나 AGI를 먼저 달성하기 위한 경쟁적인 경쟁은 안전보다는 능력에 모든 노력을 집중하도록 부추겨 "경쟁에서 승리"하도록 만듭니다.

현재 AI와 관련된 위험은 주로 기술의 인간 오용에서 비롯되었습니다. 그러나 AGI는 인간 주도의 오용을 넘어선 잠재적 위협을 제시하므로 근본적으로 다른 위험을 안고 있습니다. AGI는 자율적으로 파국적인 결과를 초래하는 계획을 생성하고 실행하여, 새롭게 발생하는 위험과 전례 없는 혼란을 인식, 분석 및 대응하는 인간의 능력을 능가할 수 있습니다. 최근 관찰된 첨단 AI의 자기 보존 경향과 결합될 경우, AGI가 통제 불능 상태에 빠지는 상황으로 이어질 수 있습니다.

이는 공동의 글로벌 관심사가 되어야 합니다. AGI 관련 위험은 특정 산업이나 사회에 국한되지 않고, 그 기원과 관계없이 전 세계적인 영향을 미칩니다. AGI의 안전하고 조화로운 통합을 위해서는 국가적 또는 기업적 노력뿐만 아니라 유엔이 주도하는 선제적인 국제 거버넌스가 필요합니다. 유엔은 위험 및 완화 전략에 대한 과학적 합의를 촉진하고, 위험 완화에 대한 공동 접근 방식에 대한 정치적 합의를 구축하며, 정책을 조율하고, 표준 또는 안전 장치를 촉진하며, 비상 사태에 대응하고, 잠재적으로 공동 안전 또는 보안 연구를 수행하거나 조율할 수 있는 독보적인 자격을 갖추고 있습니다.

글로벌 거버넌스가 없다면 글로벌 과제를 해결하기 위한 AGI의 혁신적인 잠재력이 제대로 활용되지 못하거나 잘못된 방향으로 흘러갈 수 있습니다. 더욱이 AGI가 야기할 것으로 예상되는 글로벌 재앙적 위험을 관리하는 데 글로벌 협력이 필수적일 것입니다. 유엔의 적극적인 리더십 없이는 이러한 협력이 글로벌 차원에서 달성될 수 있다고 상상하기 어렵습니다.

I. AGI 거버넌스에 대한 유엔 총회 조치의 긴급성 및 조치 없을 시 예상되는 결과

복잡한 지정학적 환경 속에서, 그리고 응집력 있고 구속력 있는 국제 규범이 없는 상황에서, 적절한 안전 조치 없이 AGI를 개발하려는 경쟁적인 움직임은 사고 또는 오용, 무기화, 그리고 실존적

실패의 위험을 증가시키고 있습니다. 국가와 기업은 안보보다 속도를 우선시하여 국가 거버넌스 체계를 약화시키고 경제적 또는 군사적 이점보다 안전 프로토콜을 부차적인 것으로 만들고 있습니다. 정부와 기업의 여러 형태의 AGI가 금세기 말 이전에 출현할 수 있고, 국가 및 국제 거버넌스 시스템을 구축하는 데 수년이 걸릴 것이기 때문에, 다음과 같은 결과를 방지하기 위한 필요한 절차를 긴급히 시작해야 합니다.

- 돌이킬 수 없는 결과:** AGI가 달성되면 그 영향은 돌이킬 수 없을 수 있습니다. 이미 많은 첨단 AI 형태가 기만적이고 자기 보존적인 행동을 보이고 있으며, 인프라와 통합된 더욱 자율적이고 상호 작용하며 스스로 개선하는 AI를 향한 추세로 인해 AGI의 영향과 궤적은 통제 불능 상태로 끝날 가능성이 있습니다. 그렇게 되면 신뢰할 수 있는 인간의 감독 상태로 되돌아갈 방법이 없을 수 있습니다. AGI가 우리의 레드 라인을 넘지 않도록 선제적인 거버넌스가 필수적이며, 이는 인간의 통제로 되돌아갈 명확한 방법 없이 통제 불능 시스템으로 이어지지 않도록 하는 것입니다.
- 대량 살상 무기:** AGI는 일부 국가와 악의적인 비국가 행위자들이 화학, 생물학, 방사능 및 핵무기를 개발할 수 있도록 할 수 있습니다. 더욱이 AGI가 통제하는 대규모의 치명적인 자율 무기 군집 자체가 새로운 종류의 대량 살상 무기를 구성할 수 있습니다.
- 중요 인프라 취약성:** 에너지망, 금융 시스템, 운송 네트워크, 통신 인프라 및 의료 시스템과 같은 중요한 국가 시스템이 AGI의 도움을 받거나 AGI에 의해 시작된 강력한 사이버 공격의 대상이 될 수 있습니다. 국가적 억제력과 국제적 협력이 없다면 테러리스트부터 초국가적 조직 범죄에 이르기까지 악의적인 비국가 행위자들이 대규모 공격을 감행할 수 있습니다.
- 권력 집중, 글로벌 불평등 및 불안정:** 통제되지 않은 AGI 개발 및 사용은 전례 없는 규모로 부와 권력의 불균형을 악화시킬 수 있습니다. AGI가 소수의 국가, 기업 또는 엘리트 집단의 손에 남아 있다면 경제적 지배력을 강화하고 지능, 혁신 및 산업 생산에 대한 글로벌 독점을 초래할 수 있습니다. 이는 대규모 실업, 법적 기반에 영향을 미치는 광범위한 권한 박탈, 사생활 침해, 그리고 제도, 과학적 지식 및 거버넌스에 대한 신뢰 붕괴로 이어질 수 있습니다. 설득, 조작 및 AI 생성 선전을 통해 민주적 제도를 약화시키고, 체계적 취약성을 증가시키는 방식으로 지정학적 불안정을 고조시킬 수 있습니다. 협력 부족은 AGI 지원, 능력 또는 통제에 대한 갈등을 초래하여 잠재적으로 전쟁으로 확대될 수 있습니다. AGI는 기존 법적 틀에 부담을 줄 것입니다. 지적 재산, 책임, 인권 및 주권에 대한 많은 새롭고 복잡한 문제가 국내 및 국제 법률 시스템을 압도할 수 있습니다.
- 실존적 위협:** AGI는 대규모 피해를 초래하거나 인간 가치와 어긋나는 방식으로 개발될 수 있으며, 심지어 현재 첨단 AI에서 이미 관찰된 자기 보존 목표에 따라 자체 목표를 진화

시켜 인간의 감독 없이 자율적으로 행동할 수도 있습니다. AGI는 인간의 개입에 관계없이 스스로 결정한 목표를 실행할 수 있도록 권력을 추구할 수도 있습니다. 각국 정부, 주요 전문가 및 AGI 개발 기업 모두 이러한 추세가 AGI 시스템이 인간을 제압하려는 시나리오로 이어질 수 있다고 밝혔습니다. 이는 먼 미래에 대한 허황된 공상 과학 소설이 아니라, 많은 주요 전문가들이 이러한 위험이 모두 금세기 안에 현실화될 수 있다고 믿으며, 그 전조가 이미 나타나고 있습니다. 더욱이 주요 AI 개발자들은 이러한 위험을 높은 신뢰도로 방지할 수 있는 실행 가능한 제안을 아직까지 제시하지 못하고 있습니다.

6. **전 인류를 위한 엄청난 미래 혜택의 상실:** 적절하게 관리된 AGI는 맞춤형 의학, 암 치료 및 세포 재생부터 개인 맞춤형 학습 시스템, 빈곤 종식, 기후 변화 해결 및 상상할 수 없는 혜택을 가져다줄 과학적 발견 가속화에 이르기까지 모든 분야, 모든 사람들에게 발전을 약속합니다. 모두를 위한 그러한 웅장한 미래를 보장하려면 위험과 혜택 모두에 대한 글로벌 인식 개선에서 시작되는 글로벌 거버넌스가 필요합니다. 유엔은 이 임무에 매우 중요합니다.

II. AGI로의 전환에 대한 유엔 거버넌스의 목적

AGI가 금세기 안에 개발될 가능성이 높다는 점을 감안할 때, 그것이 가져올 수 있는 엄청난 혜택과 엄청난 위험 모두에 대비하기 위해 강력한 거버넌스 구조를 구축하는 것은 과학적으로나 윤리적으로 필수적입니다.

AGI로의 전환에 대한 유엔 거버넌스의 목적은 AGI 개발 및 사용이 글로벌 인간 가치, 안보 및 개발과 일치하도록 보장하는 것입니다. 이는 다음을 포함합니다. 1) 점점 더 유능한 AI 시스템을 조종 및/또는 통제하기 위한 기술적 방법을 식별하기 위한 AI 정렬 및 통제 연구 발전; 2) AGI 개발 지침 제공 – AGI가 책임감 있게, 강력한 보안 조치, 투명성 및 인간 가치와 일치하여 개발되도록 보장하는 프레임워크 구축; 3) AGI 배포 및 사용을 위한 거버넌스 프레임워크 개발 – 오용 방지, 공평한 접근 보장, 위험을 최소화하면서 인류를 위한 혜택 극대화; 4) 유익한 AGI의 미래 비전 육성 – 사회, 환경 및 경제 발전을 위한 새로운 프레임워크; 5) 국제 협력을 위한 중립적이고 포괄적인 플랫폼 제공 – 글로벌 표준 설정, 국제 법률 프레임워크 구축 및 준수 인센티브 생성; 이를 통해 국가 간 신뢰를 조성하여 AGI 혜택에 대한 글로벌 접근을 보장합니다.

III. AGI에 관한 유엔 총회 세션 주요 고려 사항

AGI 거버넌스의 가장 큰 과제 중 하나는 미래 기술 개발에 대한 불확실성입니다. 이로 인해 잠재적인 혜택과 위험을 정확하게 예측하기 어렵습니다. 따라서 잠재적 혜택을 강화하면서 예상 가능한 위험을 예측하고 완화하기 위한 광범위하고 포괄적인 대응 프레임워크를 마련해야 합니다. 유

엔은 AGI 개발 및 사용에 중요한 국제적 조율을 제공할 수 있습니다. 이 과정에 모든 국가가 대표되고 지정학적 분열을 줄이는 것이 특히 중요합니다. 현재로서는 유엔만이 이러한 역할을 수행할 수 있는 적절한 위치에 있는 것으로 보입니다. AGI에 관한 특별 유엔 총회 세션에서 다음 사항을 고려해야 합니다.

A. 글로벌 AGI 관측소

AGI 관련 연구 개발의 진행 상황을 추적하고 회원국에 AI 보안에 대한 조기 경고를 제공하기 위한 글로벌 AGI 관측소가 필요합니다. 이 관측소는 글로벌 디지털 협약에서 창설된 AI에 관한 독립 국제 과학 패널 및 유네스코 준비도 평가 방법론과 같은 다른 유엔 노력의 전문 지식을 활용해야 합니다.

B. 안전하고 신뢰할 수 있는 AGI를 위한 국제 모범 사례 및 인증 시스템

AGI가 금세기 안에 개발될 가능성이 높다는 점을 감안할 때, 그것이 가져올 수 있는 엄청난 혜택과 엄청난 위험 모두에 대비하기 위해 강력한 거버넌스 구조를 구축하는 것은 과학적으로나 윤리적으로 필수적입니다.

C. 유엔 AGI 기본 협약

AGI 위험을 관리하고 공평한 글로벌 혜택 분배를 보장하기 위한 공유 목표와 유연한 프로토콜을 설정하기 위해 AGI에 관한 기본 협약이 필요합니다. 이는 표준 설정 및 라이선스 제도에서부터 고위험 AGI를 위한 공동 연구 시설, 그리고 AGI 개발에 대한 레드 라인 또는 트리거 와이어에 이르기까지 비례적인 국제적 조치를 요구하는 명확한 위험 계층을 정의해야 합니다. 협약은 글로벌 위험을 최소화하고 AGI로부터의 글로벌 번영을 극대화하는 데 필수적인, 전 세계적으로 합법적이고 포괄적이며 효과적인 AGI 거버넌스를 위한 적응 가능한 제도적 기반을 제공할 것입니다.

D. 유엔 AGI 기구 타당성 조사

AGI에 대비하기 위해 필요한 광범위한 조치와 문제의 긴급성을 감안할 때, 이상적으로는 신속한 절차를 통해 유엔 AGI 기구의 타당성을 조사하기 위한 조치가 필요합니다. IAEA와 같은 것이 제안되었는데, AGI 거버넌스가 핵 에너지보다 훨씬 더 복잡하다는 점을 이해하고, 따라서 그러한 타당성 조사에서 고유한 고려 사항이 필요합니다.

IV. 이러한 권고 사항은 미래를 위한 유엔 협약 및 기타 유엔 이니셔티브의 이행에 기여합니다.

다수의 유엔 이니셔티브는 안전하고 신뢰할 수 있는 AI 개발을 촉구합니다. 이 중 유엔 총회 AI

결의안(A/78/L.49, A/78/L.86 및 A/C.1/79/L.43)과 미래를 위한 유엔 협약, 글로벌 디지털 협약, 그리고 유네스코 AI 윤리 권고는 글로벌 위험을 선제적으로 관리하면서 모든 인류를 위한 유익한 AI 개발을 위한 국제 협력을 촉구합니다. 이러한 이니셔티브는 현재 형태의 AI에 대한 전 세계의 관심을 불러일으켰습니다. 이 보고서는 가까운 미래의 AGI 개발을 구체적으로 다룸으로써 이러한 유엔 이니셔티브를 기반으로 합니다.

미래를 위한 협약에서 이루어진 약속은 이 보고서에서 여러 방식으로 발전됩니다. AGI에 초점을 맞춘 유엔 총회 세션은 AI 거버넌스에 대한 글로벌 대화에 대한 미래를 위한 협약의 약속에 부응합니다. 유엔 AGI 기본 협약 및 유엔 AGI 기구 타당성 조사에 대한 이 보고서의 권고 사항. 우리가 제시한 관측소는 글로벌 디지털 협약의 주요 결과 중 하나인 곧 출범할 AI에 관한 독립 국제 과학 패널의 작업을 지원할 것입니다. 마지막으로 안전하고 신뢰할 수 있는 AGI를 위한 국제 모범 사례 및 인증 시스템은 유엔 총회 결의안, 유네스코 및 미래를 위한 협약에서 요구하는 신뢰와 투명성에 기여할 것입니다.

V. 결론

현재 형태의 AI와 구별되는 미래 AGI의 혜택과 위험에 대한 국내외 지도자들의 인식을 높이는 것이 시급히 필요합니다. 유엔 총회는 그러한 글로벌 논의를 시작하기에 적절한 장소입니다.

AGI의 엄청난 혜택을 누리면서 인권과 안보를 보호하기 위해서는 AGI 개발 및 사용에 대한 국제적 조율이 필요할 것입니다. 이 AGI 패널은 유엔 총회가 AGI를 위한 글로벌 거버넌스 프레임워크를 위한 특별 총회 세션에서 이러한 문제를 긴급히 다루도록 강력히 권고합니다. 그러한 조치가 없다면 극적으로 증가한 글로벌 불평등에서 실존적 위협에 이르기까지 통제되지 않은 AGI 개발 및 사용의 위험은 엄청납니다. 글로벌 관측소, 국제 인증, AGI 유엔 협약 및 전담 AGI 기구를 포함하는 이러한 유엔 주도의 접근 방식은 AGI가 위험을 최소화하면서 모든 인류에게 혜택을 주는 방식으로 개발되고 사용될 가능성을 높입니다. 이 프레임워크는 국가 간 신뢰와 협력을 증진하기 위해 포괄적이고 투명하며 시행 가능해야 합니다.

부록

위임 사항: 유엔 총회 의장단(UNCPGA) 인공 일반 지능(AGI) 고위급 패널

배경

2024년 UNCPGA 서울 선언은 인공 일반 지능(AGI) 전문가 패널을 구성하여 유엔 총회가 인공 일반 지능(AGI)으로의 전환이라는 긴급한 문제를 다루는 데 고려할 프레임워크와 지침을 제공하도

록 요청합니다.

이 작업은 유네스코, OECD, G20, G7, 글로벌 AI 파트너십 및 블레츨리 선언의 AI 가치 및 원칙에 대한 광범위한 노력과 유엔 사무총장의 AI 고위급 자문위원회, 유엔 글로벌 디지털 협약, 국제 AI 안전 연구소 네트워크, 유럽 평의회의 AI 기본 협약 및 두 건의 유엔 총회 AI 결의안의 권고 사항을 기반으로 하고 중복을 피해야 합니다. 이러한 노력은 더 좁은 형태의 AI에 더 초점을 맞춰왔습니다. 현재 AGI에 대한 유사한 관심은 부족합니다.

오늘날 세계에 잘 알려져 있고 자주 사용되는 AI와 달리 AGI는 아직 존재하지 않습니다. 많은 AGI 전문가들은 AGI가 1~5년 안에 달성될 수 있으며, 결국 우리의 통제를 넘어선 인공 지능으로 발전할 수 있다고 믿습니다. AGI에 대한 보편적으로 인정되는 정의는 없지만, 대부분의 AGI 전문가는 AGI가 인간 능력과 유사하거나 그 이상의 새로운 솔루션으로 많은 새로운 문제를 해결하기 위해 학습하고, 코드를 편집하고, 자율적으로 행동할 수 있는 범용 AI일 것이라는 데 동의합니다. 현재 AI는 이러한 능력을 갖추고 있지 않지만, 기술 발전의 궤적은 분명히 그 방향을 가리키고 있습니다.

유엔 글로벌 디지털 협약은 유엔 내에서 AI 거버넌스에 대한 글로벌 대화를 촉구합니다. AGI 민간 부문 전문가들은 AGI의 기회와 위험을 더 잘 이해하기 위한 글로벌 대화의 긴급한 필요성을 강조했습니다. AGI에 관한 유엔 총회 특별 세션은 그러한 대화를 촉진하는 가장 빠르고, 비용 효율적이며, 영향력이 가장 빠른 방법일 가능성이 높습니다.

목적

2024년 UNCPGA 서울 선언에 따라 UNCPGA 의장 및 회원들을 위해 2025년 4월 8~10일 브라티슬라바에서 열리는 UNCPGA 회의에 제출할 초기 보고서를 작성합니다. 보고서는 AGI의 위험, 위협 및 기회를 식별해야 합니다. 유엔 총회가 AGI 거버넌스를 보다 체계적인 방식으로 다루도록 인식을 높이고 동원하는 데 초점을 맞춰야 합니다. 현재의 더 좁은 형태의 AI 시스템이 아닌 아직 달성되지 않은 AGI에 초점을 맞춰야 합니다. AGI의 급속한 발전을 고려할 때 가능한 한 빨리 AGI 문제를 해결해야 하는 긴급성을 강조해야 하며, 이는 인류에게 엄청난 혜택뿐만 아니라 심각한 위험을 초래할 수 있습니다.

보고서는 이러한 전례 없는 과제를 해결하기 위한 다자간 협정과 민간 부문 조치를 모두 포함해야 합니다. 이는 인류가 직면한 가장 어려운 관리 과제가 될 수 있는 것에 대한 국제 협력 및 다자간 조치를 촉구하는 민간 부문 AGI 리더들의 요청에 응답해야 합니다.

절차

- 인류에 대한 AGI의 잠재적 위협과 AGI가 인류에게 가져다줄 수 있는 기회 및 관련 정책

문제에 대한 국제 AGI 전문가 고위급 패널(5~8명)을 소집합니다.

- AGI 패널은 2025년 1월부터 정기적으로 가상 회의를 개최하고 2025년 봄 브라티슬라바에서 열리는 다음 UNCPGA 회의를 위한 초기 보고서를 완료합니다.
- 브라티슬라바에서 열리는 UNCPGA 회의 기간 동안 초기 보고서에 대한 피드백을 바탕으로 패널은 보고서를 최종 확정하여 UNCPGA 사무총장에게 제출합니다. UNCPGA 의장의 승인을 받으면 2025년 5월 1일까지 잠정적으로 유엔 총회 의장에게 전달될 것입니다.

유엔 총회 의장단 인공 일반 지능 고위급 독립 패널 위원

- **제롬 글렌 (미국) 의장**
 - IEEE AI 조직 거버넌스 투표 회원; 유럽 연합의 Horizon 2025-27 AGI 보고서: 문제와 기회 저자; 밀레니엄 프로젝트 CEO 및 인공 협소 지능에서 AGI로의 전환의 국제 거버넌스 문제, AGI의 글로벌 거버넌스 요구 사항 및 Work/Technology 2050: 시나리오 및 행동 저자. Future Mind: Artificial Intelligence (1989) 저자.
- **헤난 아라우조 (브라질)**
 - AGI 개발 관련 위험 관리에 초점을 맞춘 AI 정책 및 전략 연구소 연구 관리자. 현재 국제 AGI 거버넌스에 대한 IAPS의 작업을 이끌고 있습니다. 옥스퍼드 중국 정책 연구소 연구원, 변호사, 콘도르 이니셔티브(브라질 학생들을 세계적 수준의 AI 연구 및 정책 기회와 연결) 공동 창립자, Rethink Priorities 및 법률 및 AI 연구소에서 AI 거버넌스 프로그램 근무.
- **요슈아 벤지오 (캐나다)**
 - 몬트리올 대학교 컴퓨터 과학 교수; 캐나다 정부 안전 및 보안 AI 자문 그룹 의장; 유엔, OECD 및 EU를 포함한 30개국이 위임한 국제 AI 안전 보고서 의장. 퀘벡 AI 연구소 Mila 과학 책임자; 과학 기술 혁신에 관한 유엔 사무총장 과학 자문위원회 위원; 튜링상 수상자 및 현재 세계에서 가장 많이 인용된 컴퓨터 과학자.
- **곽준호 (대한민국)**
 - 한국 AI 안전 연구소 기술 자문; OECD 신뢰할 수 있는 AI 개발 지침 개발에 주도적인 역할 수행; G7 하로시마 프로세스, 파리 AI 액션 서밋 준비, 한미 AI 워킹 그룹 참여, 국제 AI 안전 연구소 네트워크 한국 대표단 회원.
- **쉐란 (중국)**

- 국가 AI 거버넌스 전문가위원회 위원장; 칭화대학교 AI 국제 거버넌스 연구소 소장; OECD STI 국 자문 그룹 위원; 중국 AI 안전 연구소 자문; 유엔 지속 가능한 개발 솔루션 네트워크(UNSDSN) 리더십 위원회 공동 의장; 푸단 경영 과학 공헌상 및 중국 과학 기술 과학 협회 공헌상 수상.
- **스튜어트 러셀 (영국 및 미국)**
 - 캘리포니아 대학교 버클리 캠퍼스 컴퓨터 과학 석좌 교수 및 인간 호환 AI 센터 소장; 135개국 1,500개 대학에서 사용되고 74,000회 이상 인용된 표준 AI 교과서인 인공 지능: 현대적 접근 방식 저자; OECD AI 미래 전문가 그룹 및 세계 경제 포럼 글로벌 AI 위원회 공동 의장.
- **얀 탈린 (에스토니아)**
 - 유엔 AI 자문위원회 위원; EC AI 고위급 전문가 그룹 위원 역임; 케임브리지 대학교 실존적 위험 연구 센터 및 미래 생명 연구소 공동 창립자(두 기관 모두 AGI 문제의 선두 주자); AI 안전 센터 이사회 멤버; AGI 안전 분야 에스토니아 투자자; Skype 및 FastTrack/Kazaa 창립 엔지니어; DeepMind Google 창립 투자 이사.
- **마리아나 토도로바 (불가리아)**
 - 유네스코 AI 윤리 프레임워크 정부간 그룹 불가리아 대표; 불가리아 언론의 AGI 주요 대변인; AI 및 AGI의 윤리적 및 기술적 측면에 대한 국제적으로 인정받는 저자 및 강사; 전 국회의원 및 불가리아 대통령 자문.
- **호세 하이메 비야로보스 (코스타리카)**
 - 미래 생명 연구소 다자간 거버넌스 리드; 국제 거버넌스 혁신 센터 선임 연구원; 옥스퍼드 마틴 AI 거버넌스 이니셔티브 연구원; 법률 및 AI 연구소 연구원; 국제법 박사; 국제 AI 거버넌스에 관한 주요 서적 및 논문 공동 저자.