KISTEP 브리프 I 정책

인공지능 시대, 인간의 존엄성을 고민하는 교황청 회칙 「고귀한 인류」의 주요 내용

저자 | KISTEP 정책협력팀 배용국·IBS 고영욱

KISTEP 브리프

인공지능 시대, 인간의 존엄성을 고민하는 교황청 회칙 「고귀한 인류」의 주요 내용

2026.9.3. KISTEP 정책협력팀 배용국 팀장, 기초과학연구원 고영욱 선임정책연구원
자문: 가톨릭대학교 신학대학 조동원 부교수

요약문

□ 교황청은 2020년 「Rome Call for AI Ethics」, 2024년 「Hiroshima Addendum」, 2025년 「Antiqua et Nova」 등 인공지능 관련 논의를 지속

□ 교황 레오 14세는 '26년 5월 첫 회칙 「고귀한 인류(Magnifica Humanitas)」를 반포하며, 인공지능(AI)·로봇·디지털 전환의 시대에 인간의 존엄성에 대한 고민을 제시

- ㅇ 기술을 적대시하지 않으면서도, 효율 · 통제 · 이윤이라는 기술관료적 패러다임에 지배될 때 인간이 도구화됨을 경고하고, '바벨탑'이 아닌 '예루살렘 재건' 협력 모델을 대안으로 제시
- ㅇ 특허 · 알고리즘 · 데이터 등을 보편적 재화에 포함하는 점과, AI 거버넌스에 투명성 · 책임성 · 참여 · 보조성 강조를 요구하는 한편, 노동의 존엄, 상업화로부터의 자유, 격차 해소를 강조
- ㅇ 인공지능을 도덕적으로 중립적인 도구로 보지 않으며, 기술 혁신을 이끌고 그 사용과 한계를 정하는 최종 주체는 끝까지 사람의 양심과 판단이어야 한다고 주장

□ 동 회칙은 종교계의 문헌이나, 인공지능 기술개발과 위험성 논의에 집중하고 있는 현 시점에서 과학기술 · 인공지능 정책의 향후 방향성을 점검하는 기준으로써 시사점이 있으며, 본 고는 이러한 논의가 국내 과학기술계에서 보다 활발히 이루어질 수 있는 계기로 작용할 것을 기대

- ㅇ 인공지능 모델의 책임 있는 개발과 활용, 인공지능 · 디지털 격차 및 노동 전환 대응, 데이터 · 알고리즘의 공공성 확보 등 사회적 의제에 대하여 과학기술계 종사자의 관심 필요

[참고: 회칙(encyclical)]

전 세계 교회를 대상으로 하는 교시의 회람 서한. 교황이 교회 주교들이나 성직자들을 대상으로 반포하며, 내용 면에 있어서 교의, 윤리, 사회 문제 등을 주로 다룸(한국가톨릭대사전)

1 개요

□ 교황 레오 14세는 '25년 5월 선출된 이후 첫 번째 회칙(encyclical)으로 「고귀한 인류(Magnifica Humanitas)」를 「새로운 사태」 반포 135주년인 '26년 5월 15일 서명, 25일 반포

❍ 동 회칙은 1891년 레오 13세의 「새로운 사태(Rerum Novarum)」로 시작된 사회교리*의 전통을 잇는 문헌으로, 인공지능이 공동선을 위한 방향으로 자리잡을 수 있도록 교회가 논의에 참여

* 교회가 시대마다의 사회 문제에 신앙의 관점에서 응답해 온 가르침의 총체. 그대로 적용하는 규범집이 아니라 시대의 물음을 함께 판단하는 과정으로 규정 (『간추린 사회교리』, 2006)

❍ 회칙은 서론과 5개 장, 결론으로 구성되며, '바벨탑'과 '예루살렘 재건'이라는 두 성서적 이미지를 축으로 ①복음에 충실한 역동적 접근, ②교회의 사회교리의 기초와 원리, ③기술과 지배, ④변혁의 시대에 인간성을 보호(진리·노동·자유), ⑤권력의 문화와 사랑의 문명을 순차적으로 전개

□ 회칙은 생성형 AI의 급속한 확산, 자율무기 등 신기술 위험의 현실화, 디지털 격차 심화 등의 변화 속에서 누적된 논의를 바탕으로 마련

❍ 교황청은 2020년 「인공지능 윤리를 위한 로마의 호소(Rome Call for AI Ethics)」를 시작으로, '25년 1월 문헌(Note) 「옛것과 새것(Antiqua et Nova)」을 발표하는 등 AI 이슈를 지속적으로 다루어 왔으며, 본 회칙은 그 연장선 상에 위치

❍ 특히 「인공지능 윤리를 위한 로마의 호소(Rome Call for AI Ethics)」는 '24년 히로시마 회의에서 유대교·이슬람교·불교·힌두교 등 11개 세계 종교로 서명이 확대되며, AI 윤리가 특정 종교를 넘어선 범종교적 의제로 부상

□ 본 고에서는 과학기술계의 관점에서 「Magnifica Humanitas」의 주요 내용과 인공지능에 관한 종교계의 기존 논의를 검토하고, 과학기술 · 인공지능 정책에 대한 시사점을 도출하고자 하였음

2 회칙 「고귀한 인류(Magnifica Humanitas)」의 주요 내용

□ 인공지능 시대라는 새로운 환경을 맞이하여 '바벨탑'과 '예루살렘 재건'이라는 두 성서적 이미지를 예시로 사용자에 따라 달라지는 기술을 다루는 방향성을 제시

❍ '바벨탑' 사례는 하나의 언어, 하나의 기술, 하나의 방향만을 강요하는 획일성과 교만, 자기충족의 기획이 인간의 존엄성을 저해함을 보여줌

❍ 반면 '예루살렘 성벽 재건' 일화는 서로 다른 이들이 각자의 자리에서 참여하는 협력을 통해 인간의 존엄성, 다양성, 공동 책임이 실현될 수 있음을 강조

❍ 두 이미지는 이후 5개 장에서 기술을 판단하는 축으로 반복 등장하며, 회칙은 '무엇을 만들 수 있는가'가 아니라 '누구를 위해, 누구와 함께 만드는가'를 기준으로 삼음

〈표 1〉 성경 내 2개 사례의 비교

구분	바벨탑	예루살렘 재건
동인	이윤의 우상화, 효율 중심	공동선 · 인간 존엄 강조
다양성	획일화	다원성을 자원으로 전환
방식	일방적 Top-down	경청, 대화, 공동책임(보조성)
인간관	인간을 데이터 · 성과로 전환	인간을 목적 그 자체로 존중

□ (제1장) 복음에 충실한 역동적 접근(A DYNAMIC APPROACH FAITHFUL TO THE GOSPEL)

❍ 인공지능 확산에 따른 사회교리 판단 기준의 발전 필요성과 학계와의 열린 대화 추구(n. 23)

- 인공지능의 발전은 사회교리의 판단 기준을 근본적으로 흔드는 사안
- 기존의 판단 기준을 그대로 적용하는 데 그치지 않고, 복음에 충실하게 그 기준을 다시 검토하고 더 깊이 발전시켜야 한다고 봄
- 이를 위해 철학을 포함한 인문사회계열과 과학기술계의 기여가 반드시 필요하다고 보고, 진리와 선을 찾는 이들을 '귀중한 동맹자'로 부름

□ (제2장) 교회의 사회교리의 기초와 원리(FOUNDATIONS AND PRINCIPLES OF THE SOCIAL DOCTRINE OF THE CHURCH)

❍ 사회교리의 다섯 원리를 제시하며 이를 인공지능 시대의 판단 기준으로 삼고, 그 다섯 원리는 서로 맞물려 있어 따로 떼어 적용할 수 없다고 강조

① 공동선: 저마다 제 삶을 온전히 이룰 수 있게 하는 사회적 조건 전체
② 재화의 보편적 목적: 땅과 자원이 본래 모두의 것이라는 원칙
③ 보조성: 결정을 당사자가 내려야 하고 국가 등 상위 조직은 이를 대신하지 않고 뒷받침해야 한다는 원칙
④ 연대: 각자의 미래가 모두의 미래에 묶여 있다는 인식
⑤ 사회정의: 가장 약한 이들부터 존엄하게 살 수 있도록 사회 구조를 짜야 한다는 요구

❍ 인공지능에 대한 우려, 국가와 초국가 기구에 의한 거대 기술기업의 감독 필요(nn. 67-80)

- 특허·알고리즘·디지털플랫폼·기술기반시설·데이터도 모두를 위한 재화에 포함된다고 보고, 공유와 접근장치 없이 소수에게 집중되면 디지털 전환에 참여하는 쪽과 밀려나는 쪽의 격차 확대 지적
- 디지털 환경에서는 국가보다는 거대 기술기업이 사람들의 일상을 좌우하기 때문에, 그들에 대한 외부 검증, 알고리즘 투명성, 데이터 접근의 공정성, 구제 수단(이의 제기·피해 구제)을 요구

- 더 나아가 지역 공동체와 학교·대학, 종교 기관이 고용·서비스 접근·데이터 관리에 관한 결정에 발언권을 갖도록 국가와 초국가 기구가 공정한 규칙과 안전장치를 마련해야 한다고 주장하며, 데이터와 기술의 사용을 공적인 감독 아래 둘 것을 주문

□ (제3장) 기술과 지배. 인공지능의 약속에 비추어 본 인류의 위대함(TECHNOLOGY AND DOMINANCE. THE GRANDEUR OF HUMANITY IN LIGHT OF THE PROMISES OF AI)

❍ 인류 진보와 착취라는 기술 진보의 양면성, 혁신의 중심에는 인간 지성이 있어야 함(nn. 92-97)

- 인공지능/로봇공학/생명공학의 발전과 융합으로, 효율·통제·이윤의 논리만으로 결정을 내리는 '기술관료적 패러다임(technocratic paradigm)'이 확산되었으며, 자연은 착취대상으로, 인간은 효율적 체계의 부속품으로 취급하는 경향이 강화
- 이러한 기술적 혁신들은 인류 발전에 크게 기여할 잠재성이 있다는 점을 무시할 수 없으며, 기술적 발전에 발맞춘 윤리적·사회적 진보가 함께 필요하다고 강조
- 기술 혁신을 이끌고 그 사용과 한계를 책임 있게 정하는 주체가 끝까지 양심과 자유를 지닌 인간지성이어야 함을 주장

❍ 인공지능은 '지어진' 것이 아니라 학습으로 '배양된' 것, 윤리적 이해도 함께 발전되어야 함(nn. 98-99)

- 교황은 인공지능이 인간지능의 일부 기능을 모방하는 것이나, 속도나 연산능력에서는 사람의 능력을 넘어 분명히 이익을 준다고 인정
- 인간의 학습은 시간 속에서 선택 · 잘못 · 용서를 거치며 성장하는 과정이나, 인공지능 학습은 데이터에 기초한 통계적 적응에 그칠 뿐 내적 성장이 아니며, 도덕적 양심도 부재하다고 판단
- 현재의 AI 시스템은 '지어진(built)' 것이 아닌 '배양된(cultivated)' 것으로, 내부 작동 방식에 대한 이해가 부족하여, 기술 발전과 함께 윤리적 이해도 발전되어야 한다고 주장

❍ 인공지능을 사용할 때 유의해야 할 세 가지(nn. 100-101)

① 손쉽게 답을 얻는 편리함으로 인한 인간의 창조성과 판단력 저하

② 시스템을 설계하고 훈련한 개발자들의 가치관 선반영으로 객관성 부족

③ 실제 인격과 관계 맺고 있다는 착각으로 실제 인간과의 관계 형성 의지 상실

- 대규모 언어 모델의 확장에 따른 환경 부담을 해소하기 위한 지속 가능한 기술적 해법의 개발도 필요

❍ 인공지능 개발과 사용의 투명성과 책임성 필요, 데이터는 공동의 재화(nn. 102-108)

- 모든 기술적 도구는 무엇을 측정하고 최적화할지를 통해 이미 선택과 우선순위를 담고 있으므로, 인공지능을 도덕적으로 중립적인 도구로 볼 수 없다고 규정
- 데이터와 모델에 어떤 인간관이 내재되어 있는지를 확인할 수 있는 투명성과 인간의 삶을 좌우하는 도구로서 책임성 필요

- 인공지능에 신중한 평가와 도입 속도를 늦출 필요가 있으며, 추상적인 윤리 선언이 아니라 견고한 법적 틀과 독립적 감독이 있어야 함
- 인공지능을 인간 가치에 맞추는 기술적 정렬(alignment)에 그치면 그 가치를 정한 소수의 시각이 그대로 이식되므로, 기준 자체를 공개 논의에 부쳐야 한다고 주장
- 인공지능은 이미 자본과 데이터를 가진 쪽의 힘을 키워 사회정의와 연대를 약화시킬 수 있으므로 실효성 있는 감독이 필요하며, 데이터는 수많은 사람의 기여로 만들어진 산물이므로 공동의 재화로 관리할 방안을 찾아야 한다고 봄

❍ 인공지능의 무장 해제(disarming AI), 기술에 대한 독점적 통제를 풀어야 함(nn. 110-111)

- 더 강력한 알고리즘과 더 큰 데이터셋을 추구하는 경쟁이 군사 영역을 넘어 경제의 영역까지 확산되었으며, 기술 권력이 곧 통치할 권리를 준다는 전제를 무너뜨려야 함
- 기술을 거부하자는 것이 아니라, 기술의 독점적 통제를 풀어야 한다는 뜻
- 특히 인공지능을 개발하는 이들에게, 모든 설계 선택에는 인간을 보는 시각이 담기므로 투명성과 영향을 받는 공동체에 대한 책임, 그리고 만들어지는 것이 참으로 좋은 것이 되도록 세심히 살필 책무를 함께 져야 한다고 호소

❍ 트랜스휴머니즘·포스트휴머니즘 — 인간을 넘어서야 할 대상으로 보는 시각의 위험(nn. 115-117)

- 기술로 인간의 성능과 능력을 키우려는 흐름(transhumanism)과 인간 중심주의 자체에 도전하며 인간·기계·환경의 혼성을 내다보는 흐름(posthumanism)이 일부 기술 권력의 이념적 배경을 이루고 매체를 통해 사회·경제·정치 선택에까지 영향을 미침
- 인간을 순전히 기술적으로 완성하거나 넘어서야 할 대상으로 본다면 진보의 이름으로 희생이 정당화될 수도 있으며, 그 부담은 가장 취약한 이들에게 돌아감

□ (제4장) 변혁의 시대에 인간성을 보호. 진리, 노동, 자유(SAFEGUARDING HUMANITY AT A TIME OF TRANSFORMATION. TRUTH, WORK, FREEDOM)

❍ 인공지능으로 인해 허위정보가 크게 증폭되고 민주주의가 약화됨(n. 137)

- 디지털 플랫폼과 미디어를 보유한 쪽이 사람들의 세계관을 좌우함
- 사람들이 무엇이 진리인지를 묻지 않고 당장의 유용함에 만족한다면 민주주의가 약화됨
- 진리는 힘 있는 자들이 독점할 수 있는 것이 아니며, 콘텐츠 선정 과정의 투명함, 개인정보 보호, 검증을 위한 저널리즘과 토론의 장으로써의 역할을 강조

❍ 인공지능 시대에는 아동과 청소년의 새로운 보호 장치가 필요(nn. 140-142)

- 아동과 청소년의 감독 없는 디지털 기기 및 소셜미디어 노출은 해가 되므로, 입법을 통해 연령 제한을 두고 통제 부담을 가정에 떠넘기는 대신 서비스 제공자에게 책임을 지워야 함

- 인공지능을 사용하는 법과 함께, 언제·어떤 목적으로는 쓰면 안 되는지도 가르쳐야 하며, 플랫폼과 알고리즘에 의해 조종당할 수 있는 위험을 알리고 자기와 타인의 존엄을 지키는 법을 익히게 해야 함

❍ 기술 발전으로 인해 노동의 구조가 변화하고 있음(n. 150)

- 자동화·로봇공학·인공지능의 융합이 노동의 구조 자체를 빠르게 바꾸고 있으며, 이는 생산성 향상을 약속한다는 주장이 있음
- 그러나 생산성이 높다고 더 나은 노동이 되는 것이 아니며, 노동자가 기계의 속도와 요구에 맞추게 되면 오히려 기존의 숙련이 무의미해지고 감시와 반복업무에 처하게 됨
- 이로 인해 사람의 혁신 능력이 축적되지 않으므로, 성과가 아닌 사람을 중심에 둔 노동 시스템 재설계 필요

❍ 기술 발전으로 인한 대규모 실업은 사회적 재난으로 인식되어야 함(nn. 151, 159)

- 제4차 산업혁명 국면에서는 노동자 간 양극화가 심화될 수도 있음
- 인공지능을 도입할 때, 고용유지·재훈련·노동자 참여를 위한 조치를 함께 두는 '혁신의 사회적 기준' 제안(n. 156)
- 인공지능 시대의 새로운 노동에 적응하는 비용이 개인에게만 부담되어서는 안 되며, 기업은 노동의 질과 존엄을 스스로의 성공 지표에 포함해야 함
- 아울러 80년 넘게 GDP에 묶여 온 발전 척도로는 사람과 환경의 안녕을 담지 못하고 있으므로, 이를 보완할 지표를 개발해 노동의 존엄, 공유된 번영, 불평등 축소, 환경 보호에 미치는 영향을 평가해야 한다고 제안

❍ 인공지능/로봇공학 시대에 인간의 존엄성에 경제가 기여하기 위한 세 가지 기준(n. 164)

① 데이터와 알고리즘이 개인에게 영향을 미칠 때, 개인이 그 내부 과정을 이해하고 이의를 제기할 수 있도록 해야 함

② 기술혁신으로 인해 경제적 격차가 벌어지지 않도록 필수 서비스와 기반시설에 대한 투자가 수반되어야 함

③ 부와 권력의 집중으로 인한 불평등을 조세와 사회 보호망, 산업정책으로 바로 잡아야 함

- 이들은 혁신을 가로막기 위한 것이 아니라 혁신을 인간적인 것으로 만들기 위한 것임

❍ '디지털 주의 경제'와 데이터를 활용한 새로운 식민주의의 위협(nn. 170-179)

- 이용자의 시간과 주의를 사로잡고 취약함을 이용하도록 설계된 디지털 주의 경제(digital attention economy)에 대한 제한과 투명성 요구가 필요
- 대규모 데이터 수집으로 행동을 분류·예측하는 사회적 통제도 자유에 대한 위협이며, 무엇을 드러내고 무엇을 감출지를 정하는 방식으로 작동해 순응과 자기 검열을 부추김

- 인공지능 서비스 뒤에는 데이터 라벨링·모델 훈련·콘텐츠 조정·희토류 채굴에 종사하는 수백만 명의 노동이 있으며, 여성, 어린이, 청소년이 열악한 노동 조건에 처함
- 이에 공급망의 투명성을 확보하고 노동자 보호, 강제 노동 근절, 데이터 기반 사업 모델의 사회적 영향 평가 필요
- 원조·연구·혁신을 명분으로 취약한 지역의 건강·유전·인구 정보를 수집하고, 이를 쥔 쪽이 의약품과 투자의 배분까지 정하게 되는 새로운 식민주의를 경고하며, 데이터의 사용 주체와 목적을 정할 권한을 당사자에게 돌려주어야 한다고 요구

□ (제5장) 권력의 문화와 사랑의 문명(THE CULTURE OF POWER AND THE CIVILIZATION OF LOVE)(nn. 183-226)

❍ 전쟁이 국제정치의 수단으로 사용되며, 인공지능이 무력 사용의 문턱을 낮추고 있음

- 기계가 사람보다 일관되게 옳고 그름을 가린다는 '인공 도덕 행위자(artificial moral agent)' 개념을 부정하며, 치명적·비가역적 결정을 인공지능에 맡기는 것은 허용될 수 없음. 다만 시스템에 인간적 가치와 판단기준을 추가하는 노력은 여전히 필요
- 자율무기와 관련하여 책임성 검증, 비가역적 결정을 서두르지 않을 장고의 판단, 시민 보호 등이 필요하다고 말하며, 기술 군비 경쟁을 억제할 국제적 장치가 필요함
- 사이버공간에서는 공격 주체를 식별하기 어려워 오판에 따른 과잉 대응의 위협이 큼. 이에 따라 디지털 기술에 관한 공유 규범에 합의하고 국제기구도 개혁해야 함

❍ '권력의 문화'에서 벗어나 대화와 외교가 분쟁 해결의 표준 수단이 되는 '협상의 문화'로 옮겨 가야 한다고 주장

- 다자주의 체제가 약해지면서 비례성 원칙과 민간인 보호 같은 인도법의 성취마저 과거의 유물처럼 취급되고 있으며, 전쟁을 인간 본성의 불가피한 일로 여기는 '정치적 현실주의(realpolitik)'가 퍼지고 있다고 진단
- 과학자·기업·투자자·학계·정치인이 자기 부문만 보면 스스로 도덕적으로 중립적인 일을 한다고 여기기 쉬우므로, 전체적인 맥락에서 기술 발전을 조망할 것을 주문

3 인공지능 윤리에 대한 기존 논의

□ (Rome Call for AI Ethics) AI에 대한 윤리적 접근을 촉진하고자 교황청 생명학술원(Pontifical Academy for Life), Microsoft, IBM, 유엔식량농업기구(FAO), 이탈리아 혁신부가 '20년 2월 서명

❍ 기술 중심이 아니라 인간과 환경 중심의 디지털 혁신을 강조하며, 인공지능 개발과 활용에 있어 윤리, 교육, 권리를 강조하고, Algor-ethics의 6대 원칙*을 제시

* 투명성, 포용성, 책임성, 공정성, 신뢰성, 보안/프라이버시

❍ 2024년 개최된 다종교 컨퍼런스 “평화를 위한 AI 윤리”에서 불교, 힌두교, 조로아스터교, 바하이교 등 여러 종교 전통과 아브라함계 종교(그리스도교, 유대교, 이슬람교)를 아우르는 11개 세계 종교를 대표하는 지도자 16명이 새로 서명

□ (Antiqua et Nova) 교황청 신앙교리부와 문화교육부는 프란치스코 교황의 승인을 거쳐 ’25년 1월, AI의 올바른 사용 방향을 담은 문헌 「옛것과 새것」(Antiqua et Nova)을 공개

❍ 인공지능과 인간 지성 사이에서 지능의 개념을 구분하는 것으로 시작하여, 사회, 인간관계, 경제 · 노동, 보건, 교육, 허위정보, 전쟁 등에서 AI 발전이 가져오는 도전과 기회를 설명

❍ 인간에 대한 폭넓은 이해를 바탕으로 인공지능이 인간 지성의 풍요로움을 대체하는 것이 아니라 인간 지성을 보완하는 도구로만 사용되어야 한다는 점을 강조

□ EU는 ’19년 「신뢰할 수 있는 AI를 위한 윤리 지침」을 발표하고, 이를 인공지능법(AI Act)에 반영하는 등 인공지능 윤리에 대한 논의와 기술 발전에 따른 규정 개정을 지속 추진

❍ 2019년 「신뢰할 수 있는 AI를 위한 윤리 지침(Ethics guidelines for trustworthy AI)」을 통하여 신뢰성을 3축(적법성, 윤리성, 견고성)으로 정의하고, 4대 윤리원칙*과 7대 요건을 제시
* 인간 자율성 존중(Respect for human autonomy), 위해 방지(Prevention of harm), 공정성(Fairness), 설명가능성(Explicability)

❍ ’24년 제정된 AI Act는 전문 제7항을 통하여 윤리 지침을 고려한 고위험 AI 시스템에 대한 공통 규칙 수립을 권고하였으며, 제27항에서 7대 원칙*을 제시하며 행동강령 작성의 기초로 제시
* 인간의 참여 · 감독, 기술적 견고성 · 안전성, 개인정보보호 · 데이터관리, 투명성, 다양성 · 비차별성 · 공정성, 사회 · 환경 복지, 책임성

❍ ERA Forum을 중심으로 생성형 AI의 연구 · 활용에 있어 연구자, 연구기관, 연구지원기관 등에게 권고하는 가이드라인*을 개정 · 발표('26.5., 제3판, 초판 '24.3.)
* Living guidelines on the responsible use of generative AI in research

□ 과학기술정보통신부는 안전하고 신뢰할 수 있는 AI 기본사회 구현을 위해, AI를 개발·제공·이용하는 모든 주체가 함께 지켜나갈 공통의 가치로 「대한민국 인공지능 윤리원칙」 제정

❍ 우리나라는 ’25년 1월 제정된 「인공지능 발전과 신뢰 조성 등에 관한 기본법」을 통해 인공지능 윤리를 정의하고, 정부의 윤리원칙 제정 권한, 사업자의 투명성 · 안정성 확보 의무, 고영향 인공지능 영향평가 등을 법으로 규정

❍ 과기정통부는 2020년 제정한 「인공지능 윤리기준」을 계승하고, 기술 발전과 달라진 사회 여건을 반영하기 위하여 ’25년 12월부터 「대한민국 인공지능 윤리원칙」 제정을 추진, ’26.8월 발표

- 인간, 사회, 인류 차원에서 추구해야 하는 근본적 지향점인 3대 가치*와 이를 실현하기 위해 모든 주체가 실천해야 할 공통의 기준으로 7대 원칙 발표
* 인간의 존엄성, 사회의 공공선, 인류의 지속가능성

- 정부는 사례 중심의 해설서와 분야별 자율점검표 마련 · 보급, 대상별 윤리교육 교재 마련 및 확산, 기술 · 사회 변화에 따른 윤리원칙 점검 · 보완 등을 향후 계획으로 제시

〈표 2〉「대한민국 인공지능 윤리원칙」 7대 원칙의 주요 내용

원칙	주요 내용
인간중심성 (Human-centeredness)	사람의 판단을 지원하고 창의성·문제 해결 역량을 강화하는 방향으로 개발·활용, 필요한 경우 사람이 AI의 동작에 개입할 수 있도록 노력, 과의존 경계
프라이버시 보호 (Privacy Protection)	사생활 침해에 활용되지 않도록 하고, 개인정보는 정해진 목적에 한해 필요한 범위에서 처리, 정보주체의 개인정보 자기결정권 존중
공정성·포용성 (Fairness and Inclusiveness)	부당한 차별로 이어지는 편향 방지, 사회적 약자를 비롯한 모든 사람의 공평한 접근과 혜택·기회, 이해관계자 참여 보장 노력
책임성 (Accountability)	각자의 역할과 책임을 명확히 인식·이행, 문제 발생 시 책임 소재를 밝히고 실질적 피해 구제 노력, 영향력이 큰 주체는 상응하는 책임 이행
안전성 (Safety)	생명·신체·정신적 건강의 위해와 재산상 피해 방지, 위험 수준에 맞는 안전조치 마련, 오남용·외부 공격 대비
신뢰성 (Reliability)	목적과 활용 맥락에 맞는 최소 성능 기준 설정·유지, 의도된 목적과 허용된 권한 범위 안에서 작동, 성능·작동 상태 지속 점검
투명성 (Transparency)	AI 활용 사실과 수준·한계 등 필요한 정보를 알기 쉽게 제공, 판단 기준과 위험 수준의 공유·이해 노력

※ 자료: 「대한민국 인공지능 윤리원칙」('26.8.)

4 결론 및 시사점

□ 인공지능 기술의 급속한 발전이 이루어지는 가운데 종교계에서는 인류에 대한 의도적/비의도적 위협 가능성을 우려하고, 윤리와 가치, 포용성 등에 중점을 둔 인간 중심의 기술개발을 강조

❍ 회칙 「고귀한 인류(Magnifica Humanitas)」는 러다이트 운동(Luddite Movement)과 같이 기술발전을 거부하는 것이 아니라, '누구를 위해, 누구와 함께 만드는가'를 기준으로 기술발전의 방향성에 대한 논의 필요성을 제기

❍ 과학이 가치중립성을 가진다는 명제는 지속적으로 비판받고 있으나, 인공지능 기술이 매우 빠르게 진화하는 현재 종교계를 포함한 각 계의 우려가 존재

❍ 최근 수립된 「대한민국 인공지능 윤리원칙」은 인간중심성 · 공정성 · 포용성 · 책임성 등을 주요 원칙으로 포함하며 '바벨탑' 사례가 아닌 '예루살렘 재건' 사례와 방향을 같이 하고 있음

□ 「고귀한 인류(Magnifica Humanitas)」는 해외 과학계에서는 토론의 대상으로 주목을 받고 있으며 우리나라 과학기술계에서도 보다 심도 깊은 논의가 필요

❍ Nature, MIT Technology Review, Bulletin of the Atomic Scientists 등에서 기고자들은 관점에 따라 회칙의 중요성, 한계, 비판 등 다양한 논의를 진행

❍ 국내에서는 과학기술 · 인공지능 이슈를 다루는 전자신문, ZDNet Korea 등 일부 언론에서 회칙 발표를 기사로 전달하였으며, 천주교, 불교, 개신교 등 종교계에서 인공지능에 대한 논의 지속

□ 인공지능 기술의 발전과 세계 시장 선점을 위해 노력하는 동시에 인공지능을 포함한 과학기술의 사회적 가치에 대한 논의가 우리 사회에서도 지속되어야 함

❍ 인공지능 모델의 책임 있는 개발과 활용, 인공지능 · 디지털 격차 및 노동 전환 대응, 데이터 · 알고리즘의 공공성 확보 등의 사회적 의제에 대하여 과학기술계 종사자의 관심이 요구됨

❍ 과학기술정보통신부 주도로 수립 한 「대한민국 인공지능 윤리원칙」은 세계적으로도 의의가 높으며 단순한 원칙으로 머무르지 않고 사회 전반으로 확산 · 적용되기 위한 지속적 노력 필요

참고문헌

- Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education, Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence, https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.html#AI_and_Society, 2025.1.28.
- European Commission, Ethics guidelines for trustworthy AI, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, 2019.4.8.
- European Commission, Commission starts enforcing AI Act rules and new transparency requirements on 2 August, https://ec.europa.eu/commission/presscorner/detail/en/ip_26_1714, 2026.7.31.
- ERA Forum Stakeholders' document, Living guidelines on the responsible use of generative AI in research, Third version, https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf, 2026.5., (2026.8.21.조회)
- Paolo Benanti, I advise the Vatican and the UN on AI — don't dismiss the Pope's message as theology, Nature, https://www.nature.com/articles/d41586-026-01876-z, 2026.6.12.
- Pope Leo XIV, Magnifica Humanitas: On Safeguarding the Human Person in the Time of Artificial Intelligence, https://www.vatican.va/content/leo-xiv/en/encyclicals/documents/20260515-magnifica-humanitas.html, 2026.5.15.
- REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), http://data.europa.eu/eli/reg/2024/1689/oj, 2024.6.13.

- Sara Goudarzi, The church steps into the AI debate: What the Pope's first encyclical means, Bulletin of the Atomic Scientists, https://thebulletin.org/2026/06/the-church-steps-into-the-ai-debate-what-the-popes-first-encyclical-means/, 2026.6.2.
- Séamus Finnarchive & Susan Francoisarchive, How the Pope's Magnifica Humanitas offers a template for individuals to meet the AI moment, MIT Technology Review, https://www.technologyreview.com/2026/05/29/1138107/how-the-popes-magnifica-humanitas-offers-a-template-for-individuals-to-meet-the-ai-moment/, 2026.5.29.
- ZDNet Korea, 교황, AI 시대를 향한 거룩한 선전포고…위대한 인간성을 묻다, https://zdnet.co.kr/view/?no=20260526100121, 2026.5.26.
- 가톨릭신문, 「교황청, 인공지능 올바른 사용 위한 가이드라인 발표」, https://www.catholictimes.org/article/20250203500131, 2025.2.5.
- 가톨릭신문, 「옛것과 새것」(Antiqua et Nova)—인공지능의 올바른 사용을 위한 지침」, https://www.catholictimes.org/article/20250207500045, 2025.2.12.
- 과학기술정보통신부, 「과기정통부, '인공지능 윤리원칙' 초안 공개…대국민 의견수렴 착수」, https://www.msit.go.kr/bbs/view.do?sCode=user&mId=307&mPid=208&bbsSeqNo=94&nttSeqNo=3187426, 2026.5.28.
- 과학기술정보통신부, 미래기술의 씨앗을 성장동력으로첨단기술 실증·사업화와 인공지능 신뢰 기반 논의, https://www.msit.go.kr/bbs/view.do?sCode=user&mId=307&mPid=208&pageIndex=&bbsSeqNo=94&nttSeqNo=3187684&searchOpt=ALL&searchTxt=, 2026.8.24.
- 과학기술정보통신부, [보도참고] 정부, 「대한민국 인공지능 윤리원칙」 제정, https://www.msit.go.kr/bbs/view.do?sCode=user&mId=307&mPid=208&pageIndex=&bbsSeqNo=94&nttSeqNo=3187686&searchOpt=ALL&searchTxt=, 2026.8.24.
- 바티칸 뉴스, 「교황 "AI 윤리, 인류 가족의 선익 보호해야"」, https://www.vaticannews.va/ko/pope/news/2023-01/pope-francis-receives-rome-call-vatican-audience.html, 2023.1.10.
- 바티칸 뉴스, 「교황 "치명적 자율살상무기의 사용을 금지하고, 기계(머신)의 시대에 인간의 존엄성을 보호해야 합니다"」, https://www.vaticannews.va/ko/pope/news/2024-07/papa-hiroshima-ai-etica-pace-religioni-dignita-umana-tecnologia.html, 2024.7.10.
- 바티칸 뉴스, 「교황 레오 14세의 첫 회칙: "인공지능은 무장 해제되어야 합니다."」, https://www.vaticannews.va/ko/pope/news/2026-05/papa-leone-enciclica.html, 2026.5.25.
- 연합뉴스, AI에 위안 얻는 시대…인간과 AI 바람직한 공존 모색하는 종교계, https://www.yna.co.kr/view/AKR20260524029800005, 2026.5.25.
- 전자신문, 레오 14세 교황 "인류 보호 위해 AI '무장해제'해야"…첫 회칙 발표, https://www.etnews.com/20260526000006, 2026.5.26.
- 천주교 마산교구, 「옛것과 새것(Antiqua et Nova)」, https://cathms.kr/board_11/23870, 2025.7.16.
- 한국천주교주교회의, 「한국가톨릭대사전」, https://encyclopedia.catholic.or.kr/, 접속일: 2026.7.15.
- 한국천주교중앙협의회, 『간추린 사회교리』, 2006 번역·출판, 교황청 정의평화평의회 2004 작성.

저자

KISTEP 정책협력팀 배용국 팀장/연구위원 (gook@kistep.re.kr, 043-750-2415)

기초과학연구원 국제협력팀 고영욱 선임정책연구원 (young@ibs.re.kr, 042-878-8142)