Institute for Peace-Sharing

AI의 무기화, 가톨릭교회는 무엇을 할 것인가

평화나눔의 시선

2026.09.15. | 제26-05호

로마선언: 바티칸에 모인 노벨상 수상자들

지난 7월 14일부터 15일까지 바티칸에서는 'AI와 핵무기에 관한 노벨상 수상자 회의'가 열렸다. 역대 노벨상 수상자들과 교황청 고위 성직자 및 전직 국가 수반, 그리고 핵물리학자와 국제정치학자, 시민사회 지도자 등 200여 명이 한자리에 모여 핵무기와 AI의 결합이 가져올 수 있는 미래 위험에 대해 경고하고 이를 적절하게 통제해야 한다고 한목소리를 낸 것이다.

지난 7월 14일-15일 바티칸은 교황의 여름 별장인 카스텔 간돌포에서 'AI와 핵무기에 관한 노벨 수상자 회의'를 열고 '로마선언'을 채택했다.

평화상뿐만이 아니라 물리학상, 생리의학상 등 30여 명의 노벨상 수상자들은 이틀간의 회의를 마친 뒤 로마시청에서 'AI와 핵무기에 관한 로마선언'을 채택했다. 이들은 '로마선언'을 통해서 AI기술의 개발과 사용 단계에서 책임성 제고와 AI 규제 거버넌스의 확립, 핵 군축 등을 촉구했다. 특히 이 회의에는 구글 딥마인드, 오픈AI, 앤트로픽 등 주요 AI 기업의 책임자급 인사들도 참석해 관심을 끌었다.[1)]

가톨릭의 심장부인 바티칸에 노벨상 수상자들과 전문가들이 모여서 AI 기술의 무분별한 확산이 가져올 위험성을 경고하고 나선 데에는 교황 레오 14세의 강력한 의지가 작용했다고 볼 수 있다. 레오 14세 교황은 지난 5월, AI 기술의 급속한 확산 속에서도 인간의 존엄성을 지키기 위한 가톨릭교회의 지침서로서 '고귀한 인류(Magnifica Humanitas)'라는 회칙을 발표한 바 있다. 레오 14세 교황이 즉위 1년이 된 시점에 회칙이라는 권위있는 문건을 통해 AI의 무분별한 확산이 가져올 수 있는 위험성을 경고하고 나섰다는 점은 의미심장하다.

레오 14세의 회칙 '고귀한 인류'는 135년 전 교황 레오 13세(재위 1878-1902)가 회칙 '새로운 사태(Rerum Novarum)'를 통해 산업혁명 직후 노동환경의 악화로 인해 인간의 존엄성이 위협받는 상황에 경종을 울렸던 상황에 비견할 만하다. 레오 13세의 회칙이 노동권의 등장과 사회보장제도 도입이라는 흐름 속에서 근대적인 복지국가가 지향해야 할 윤리적 규범을 확립하는 계기를 제공했다면, 레오 14세의 회칙은 AI 시대의 도래라는 '새로운 사태' 앞에서 인간 노동의 위축 및 AI의 무기화가 불러올 인류의 위기에 대처하기 위한 경각심을 촉구한 것으로 볼 수 있다.

가톨릭교회가 AI의 일상화를 '새로운 사태'로 규정하는 이유는 두말할 것도 없이 인공지능의 탐구 역량과 문제 해결 방식이 하루가 다르게 진화하고 있기 때문이다. AI 기술발전의 최종 목적지가 어디일 지에 대해 자신 있게 이야기해줄 수 있는 사람은 아무도 없다. 다양한 시나리오들이 존재할 뿐이라는 점에서 우리는 지금 "상상할 수 있지만 완전히 예측할 수 없는 방향"[2)] 으로 나아가고 있는 것이다. 한 가지 분명한 것은 우리 세대보다 미래 세대에서, 미래 세대보다 그다음 세대에서 인공지능에 대한 의존도가 점점 높아질 것이라는 점이다. 그렇다면 인공지능이 인간의 지적 능력을 대체하는 것은 이미 예정된 미래일까.

1) 성기영, "AI와 핵무기의 결합 경고한 노벨상 수상자들," 『한겨레』 2026년 7월 21일 https://www.hani.co.kr/arti/opinion/because/1269130.html (검색일: 2026년 9월 1일)
2) 교황 레오 14세, "고귀한 인류", 4항.

인공지능:
인간과의 관계 맺기 방식에 대한 근본적인 성찰 불가피

가톨릭교회는 인간의 지능을 사실의 습득이나 특정한 과업의 수행능력으로 한정 짓는 방식에 단호히 반대해 왔다. 인간의 지성을 '인간 안에 있는 하느님 모습의 표현'으로 받아들이는 교회는, 존재의 충만함을 이해하고 성찰할 수 있는 인간만의 능력에 주목한다. 인공지능이 인간의 지능이 생산하는 결과물의 일부를 모방하는 비범한 기술적 성취인 것은 사실이지만, 인간의 지성이 수많은 과정을 통해서 체득한 경험을 바탕으로 내재화한 유기적 결실을 대체할 수는 없기 때문이다. 따라서 인간의 존엄성에 기반한 윤리적 식별이나 인류 공동체의 관계성 회복을 위한 최소한의 덕목을 인공지능에게 기대할 수는 없다. 무엇보다도 인공지능은 인간의 지능이 기계적으로 구현된 형태가 아니라 인간 지능이 만들어낸 산물이라는 관점을 분명히 하는 것이 필요하다.[3)]

그럼에도 불구하고 현재와 같은 AI의 발전 속도에 비춰볼 때 인간과 AI의 관계 맺기 방식에 대한 근본적 성찰은 불가피해 보인다. 우리 인류는 이미 AI가 인간의 모든 능력을 대체하는 범용인공지능(AGI)의 시대 진입을 앞두고 있다. 또 공상과학영화에서 본 것처럼 AI가 인간의 능력을 뛰어넘는 고도의 신경망 구조를 갖게 되는 초인공지능(ASI) 시대의 도래를 예측하는 이들도 적지 않다. 전문가들이 AI 기술의 발전 속도가 인간의 예측을 넘어서면서 시공간의 규칙이 근본적으로 붕괴되는 기술적 '특이점(singularity)'을 언급하는 것도 이러한 이유에서이다.

이러한 기술적 특이점을 넘어서는 순간 인류의 눈앞에는 AI 기술이 가져올 찬란한 유토피아 대신 디스토피아로 가는 어둠의 터널이 시작되지 않으리라는 보장이 없다. AI 유토피아론자들은 AI가 인류의 모든 질병과 재앙을 막아주고 사회문제를 해결해줄 수 있는 '치트키'라고 주장할 것이다. 그러나 이러한 시나리오가 설득력을 갖기 위해서는 ASI의 시대에 진입한 인류가 동물원 우리 안에서 사육되는 영장류들처럼 인공지능의 억압과 통제 속에 놓이지 않아야 한다는 전제가 충족되어야만 한다.

문제의 초점은 인간이 AI라는 도구에 대한 통제력을 지속적으로 유지할 것인가, 아니면 반대로 오히려 AI가 인류를 도구로 삼을 것인가로 모아질 것이다. 선사 이래 인간은 도구의 발명을 통해 문명의 확산과 진화를 도모해 왔다. 돌도끼로부터 종이와 인쇄술, 바퀴와 나침반 등 모든 도구의 발명이 인간의 삶을 윤택하게 한 것은 이들 도구들의 사용 주체가 다름 아닌 인간이었기 때문이다.

3) 교황청 신앙교리부·문화교육부, "옛것과 새것(Antiqua et Nova): 인공지능과 인간지성의 관계에 관한 공지" 제3장 29항-35항.

그러나 AI라는 또 하나의 도구가 등장한 이후 인간과 도구의 관계 맺기 방식은 달라졌다. 인간이 감당해야 할 지적 노력의 일부를 대신해주는 차원을 넘어 AI가 인간의 지능이 도출하는 결과와 맞먹거나 이를 뛰어넘는 결과를 생산하기 시작한 것이다. 인간은 AI가 제공하는 알고리즘 권력에 이미 장악되었고 AI가 제공하는 정서적 위안에서 가족이나 친구가 보여주지 못한 따뜻함을 느끼기까지 한다. 나아가 AI는 이제 인간의 명령에 사사건건 의지하지 않고도 독자적인 진화를 위한 재생산 메커니즘을 갖추게 되었다. 이른바 '재귀적 자기개선(Recursive self-improvement)' 능력의 획득이다.

인간의 전유물로 여겨졌던 분석과 예측, 평가와 의사결정 모두를, 그것도 인간 두뇌의 처리 능력을 훨씬 뛰어넘는 속도와 깊이로 해냄으로써 인간의 효능감을 위협하고 있다. 인류는 이제 하나의 결정에 직면해 있다. AI와 협력할 것인지, 경쟁할 것인지, 아니면 AI의 지배력에 맞서 싸워나갈 것인지. 인간의 창의적 활동은 물론 에이전트 AI와 같은 자율형 인공지능 시스템 분야에서 AI와의 협력은 필수적이다. 한편, 피지컬 AI의 확산을 통해 인간의 노동이 기계로 대체되고 청년들의 일자리가 사라지는 상황은 AI와의 경쟁을 불가피하게 만들 것이다.

AI의 무기화: 최종 공격 판단을 누가 내리는가

안보 분야에서 AI와 무기의 결합은 어떻게 진행되고 있는가. 우리는 AI가 탑재된 무기가 무장한 정예병력을 어떻게 무력화시키는지를 이미 체감한 바 있다. 2025년 7월, 우크라이나 하르키우 전선에서 AI가 탑재된 무인 드론 공격에 노출된 러시아 병사들은 아무런 저항 없이 항복 표식을 들고 벙커를 뛰쳐나왔다. 지상에서 적군과 조우 해보지도 못한 병사들은 우크라이나군이 조종하는 드론을 따라 일렬로 이동하며 포로의 대열에 합류하지 않을 수 없었다. 우크라이나 전선에 등장한 AI 드론의 조종 주체는 인간이었지만, 이들 무인 드론의 조종 주체가 우크라이나군이 아닌 AI라고 가정해보자. AI는 공포에 질려 생존을 갈구하는 병사들의 몸동작과 표정을 읽어낼 것인지, 읽어내더라도 병사들의 행동이 진심인지, 위계인지에 대한 판단 앞에서 인간의 통제를 벗어난 AI는 어떤 결정을 내릴 것인지 인간은 아직 답변을 갖고 있지 못한다.[4)]

지난 2월 미군의 이란 초등학교에 대한 오폭으로 어린이를 포함해 170여 명의 민간인이 희생되었다. 이 사건은 이란 군사시설에 대한 데이터를 기반으로 AI가 표적을 제공하고 이를 미군이 승인한 결과였다. 오폭의 원인으로는 과거 혁명수비대 기지로 쓰였던 초등학교 건물을

4) 최재운, "인간없는 전쟁" (서울: 북트리거, 2026), pp.277-279.

AI가 업데이트 하지 못한 채 표적으로 설정했다는 분석이 나오고 있지만 미국 측은 조사 결과를 공개하지 않고 있다. 뿐만 아니라 이스라엘군의 가자 공습 과정에서는 표적 데이터의 분석과 탐지에 소요되는 시간이 과거 72시간에 20분으로 줄어들었고, 이를 인간이 검토해 승인하는 데 걸린 시간은 20초였다는 통계도 등장하고 있다. 모두 AI 기술의 무기화로 인해서 발생하는 현상들이다. 이처럼 앞으로도 우리는 수많은 병사들의 목숨이 오가는 전장에서 인간 대 인간의 소통이 배제된 채 AI의 독자적 판단에 의해 살육이 자행되는 광경을 수도 없이 목도할지 모른다.

AI는 이제 현대전의 승패를 좌우하는 핵심적 기술요소가 되어가고 있다. 군비경쟁은 곧 AI 경쟁이 되었다. AI 강대국들은 허위정보의 유포를 통한 심리전에서부터 정찰, 지휘통제, 통신, 공격, 방어 등 전쟁 수행의 모든 측면에서 우위를 유지하기 위해 고도의 AI기술을 활용할 것이다.

전쟁 계획과 수행을 위한 과정에서 AI에게 인간의 의사결정을 위임하게 되면 전쟁의 양상은 판이하게 변화할 것이다. 교전 상대국의 군사 독트린에 기반한 전략과 전술에 대한 인간의 이해를 뛰어넘어, AI의 판단에 의해 정세 분석이 제공되고 선제공격 및 방어 등 의사결정이 이뤄질 경우 종래의 전쟁에서는 예상하지 못했던 위험요소들이 등장할 수 밖에 없다.[5)]

트럼프 대통령의 등장과 러시아의 우크라이나 침공 이후 핵 비확산 제제가 위협받고 있는 국면에서 AI 기반 군비 경쟁이 촉발되었다는 사실은 더욱 위험한 시나리오를 예고하고 있다. 핵무기와 AI는 모두 절대적인 힘(total power)을 가진 기술인 동시에 인류 공동체의 안위를 좌우할 수도 있는 능력을 보유하고 있다. 이처럼 절대적인 힘을 가진 두 가지의 기술이 결합될 경우를 상정하면, 핵무기의 사용에 관한 최종적 판단에서 AI가 인간의 결정을 대신할 수도 있다는 우려는 더 이상 가공의 시나리오가 아니라고 할 수 있다.

미국이 일본 히로시마와 나가사키에 인류 최초로 핵폭탄을 투하한 이후 핵보유국과 잠재적 핵개발국들 사이에서는 억지(deterrence)의 국제정치가 작동해 왔다. '상대국이 핵무기를 사용하지 못하도록 하기 위해 핵무기를 개발하고 유지한다'는 논리이다. 비록 불완전하지만 핵무기의 사용과 확산을 금지하는 국제사회의 약속도 존재하고 있다.

그러나 AI 기술과 핵무기 체계의 통합이 가져올 재앙적 결과를 예방하기 위한 구체적 논의는 제대로 이뤄지지 않고 있다. 게다가 '경보 즉시 발사(launch-on-warning)' 시스템과 같은 핵 태세를 유지하는 한 AI와 핵무기의 통합을 완벽하게 제어할 수 있을지조차 불분명한 현실이다. 국가 대 국가의 핵 대치 상황에서 상대국의 조기 경보를 교란하기 위해 AI를 사용할 경

5) 헨리 키신저·에릭 슈밋·대니얼 허튼로커, "AI 이후의 세계" (서울: 윌북, 2023), pp.202-203.

우 위협 통제의 실패로 인해 확전의 위험은 증가할 것이다. 핵 확전에 대비하기 위해서는 파괴력을 더욱 높인, 다종다양한 핵무기가 필요하기 때문에 추가적인 핵 군비 경쟁이 불가피할 수도 있다.

미래 시나리오: 인류 공영인가, 인류 존망인가

세상이 AI가 대체할 노동의 미래와 인류의 공영을 위한 기술 진보에 환호하는 사이 진보를 거듭한 AI 기술의 종착지가 결국 우리 인류의 안전을 위협하는 살상무기 개발일 수 있다는 사실은 역설적이다. 가톨릭교회는 오래전부터 기술 진보라는 과학의 성과는 외면하지 않으면서도 인류가 이룩한 기술적 성취의 희생자가 될지도 모르는 위험을 경고해 왔다. 교황 성 바오로 6세(1897~1978)가 "가장 놀라운 기술적 업적과 가장 경이로운 경제 성장이 진정한 도덕적, 사회적 진보와 동반되지 않는다면 인간을 거스르는 방향으로 작용할 것"이라고 경고한 것은 오늘날과 같은 AI시대의 도래를 예상하기조차 어려웠던 시기였다.[6]

국제사회는 이제 핵무기를 포함한 대량살상 무기의 AI화를 어떻게 규제하고 인간의 통제권을 회복해야 할 것인지에 대한 대답을 내놓아야 한다. 가톨릭교회는 AI에 대한 통제권 상실이 인간의 존엄성에 대한 실질적 위협을 초래할 것이라는 문제의식으로 이미 오래전부터 AI의 위험성을 경고해 오고 있다. 2024년 세계 평화의날(1월 1일)을 맞아 '인공지능과 평화'라는 담화를 발표했던 프란치스코 교황은 무분별한 AI 확산의 배후에 있는 기술권력의 문제점을 지적하고 인간을 위한 AI의 필요성을 강조했다.[7]

돌이켜보면 국제질서가 요동치면서 세계적 차원의 전쟁을 예고하는 일촉즉발의 순간이나 국지전을 뛰어넘는 장기적 전쟁국면으로 접어들었을 때 교회는 성서적 통찰을 기반으로 역사적 문헌을 발표해 왔다. 1963년 쿠바 미사일 위기로 인류가 3차대전 발발 위험에 직면했을 때 교황 요한 23세는 '지상의 평화' 회칙을 발표했다. 교황 레오 14세가 '고귀한 인류' 회칙을 통해 AI에 대한 인간의 통제권 상실이 초래할 위험성을 경고하고 나선 것 또한 AI의 무기화가 과거 두 차례의 세계대전이 초래한 결과 이상으로 인류의 존재 자체에 대한 위협으로 이어질 수 있다고 내다보았기 때문이다. 프란치스코 교황이 지적한 대로 '조각난 형태의 세계 3차대전'이 이

6) 교황 레오 14세, "고귀한 인류", 94항.
7) 바티칸 뉴스, "교황 'AI 개발, 소수의 기술 대기업 권력 제한하고 모든 인간에게 혜택 돌아가도록 해야'" https://www.vaticannews.va/ko/pope/news/2024-06/papa-francesco-fondazione-centesimus-annus-ia-tecnologia-etica.html. (검색일: 2026년 9월 3일)

미 시작되었다. 따라서, AI의 무기화와 이에 대한 인간의 통제권 상실 가능성을 하루빨리 차단하지 않으면 3차대전은 인류 전체를 공동의 패배자로 만들어버리는 결과를 낳을 수 있다.

지난 7월 바티칸에 모인 노벨상 수상자들이 채택한 'AI와 핵무기에 관한 로마선언'의 마지막 구절은 "우리 세대와 미래 세대가 존망의 위기에 처해 있다"고 경고하고 있다. 1981년 히로시마 평화기념공원에서 성 요한 바오로 2세 교황은 "인류가 전쟁을 끝내지 않으면, 전쟁이 인류를 끝낼 것"이라고 경고한 바 있다. 반세기 전의 경고가 현실이 되는 상황을 막기 위해서 인류 공동체는 구체적 행동에 나서야 한다. 〈끝〉