

인공지능 시대의 영성: 하나님의 형상(*Imago Dei*)으로서 인간 고유성 위협의 신학적 성찰과 교회의 목회적 대응*

Spirituality in the Age of Artificial Intelligence: A Theological Reflection on the Threat to Human Uniqueness as the *Imago Dei* and the Church's Pastoral Response

구자영 (Jayoung James Goo)**

ABSTRACT

This paper offers a theological reflection on the threat to human uniqueness posed by artificial intelligence (AI), and proposes a pastoral response for the Korean church. Adopting an interdisciplinary method connecting an engineering analysis of how AI systems learn with theological anthropology, the study first diagnoses the crisis of human uniqueness along four dimensions: individuality, creativity, empathy, and ethics. It then argues that AI cannot possess spirituality, drawing on the limits of computationalism, the irreducibility of consciousness, and neurological evidence of spiritual experience, and grounds this argument in an immanent analysis of AI's learning structure: the convergence of weights toward the statistical mean, the avoidance of overfitting required for general use, and the cost structure of retraining that renders repentance structurally impossible for AI. The central claim is that AI's limitations are not a temporary deficiency of immature technology but a structural consequence of its operating principles. On this basis, the paper reinterprets the *Imago Dei* as an event of spirituality realized in relationship with God, proposing four dimensions AI can never possess: relational spirituality, spirituality as co-creator, spirituality of vulnerability, and spirituality of the cross, and frames the age of AI as a kairos calling the church back to its essence.

Key words: artificial intelligence, spirituality, *Imago Dei*, theological anthropology, *Coram Deo*

* 2026년 8월 4일 접수, 9월 4일 최종수정, 9월 5일 게재확정
이 논문은 2026학년도 부산외국어대학교 학술연구조성비에 의해 연구되었음.

** 부산외국어대학교(Busan University of Foreign Studies) 만오교양대학 교수, 부산광역시 금정구 금샘로 485번길 65, 20256002@bufs.ac.kr

I. 서론

인류는 앨런 튜링이 반세기 전 제시했던 ‘생각하는 기계’가 현실이 된 문명사적 변곡점 위에 서 있다(Turing, 1950: 433-460). 인공지능(Artificial Intelligence, 이하 AI)은 더 이상 공상과학 소설이 아니라 인간의 사유와 관계, 창조와 윤리의 문법을 근본적으로 재편하는 실존적 도전이 되었다. 기술이 인간을 규정하는 주체가 되고 인간은 알고리즘의 선택을 받는 객체로 전락할 수 있다는 경고가 현실로 다가왔다. 이제 “인간이란 무엇인가?”라는 신학사를 관통해온 근원적이고 본질적인 물음은 현대신학의 가장 시급한 과제로 다시 부상하게 되었다.

국내 신학계의 AI 담론은 김동환(2016: 35-60)이 다학제적 연구의 필요성을 제기한 이래 꾸준히 확장되어 왔다. 김난예(2017)는 AI 시대에 요청되는 영적 민감성을 물었고, 김경식(2023)은 AI가 과연 ‘인간의 형상’이 되어가고 있는지를 검토하였다. 박종환(2024)은 강인공지능(AGI)이 하나님의 형상을 지닐 수 있는지를 정면으로 논구하였다. 서미경(2024)은 바빙크의 세계관을 토대로 생성형 AI 시대에 신앙과 학문을 통합하는 길을 모색하였다. 박영진(2026)은 판단의 주체와 책임이 사용자로부터 AI로 이동하는 ‘결정 위임’ 현상을 하나님의 주권과 자유의지라는 조직신학의 틀에서 고찰하였다. 김성민·김미경(2026)은 텍스트 마이닝을 통해 AI 노동 담론의 의미 구조를 성경적 경제관과 대비하였다.

이러한 축적에도 불구하고 하나의 공백이 반복해서 나타난다. 기존 논의는 대체로 조직신학적 인간론의 재해석에 머물거나 기술 현상에 대한 윤리적 논평에 그쳐, AI의 작동 원리에 대한 내재적 분석과 신학적 성찰이 충분히 결합되지 못하였다. 김영록(2026: 122-123)은 하나님의 형상을 기능적 이해에서 관계적 이해로 전환할 것을 설득력 있게 논증한 뒤, “자율적이고 비물질적인 ‘기술 권세’가 가지는 현대적 특수성을 사회학적·공학적으로 충분히 분별하고 해부하는 데에는 지면의 한계가 있었다”고 술회한다. 곧 AI 시대의 신학적 인간론은 ‘기능에서 관계로’라는 전환의 방향은 확보하였으나, 그 전환을 요구하는 근거를 기술 내부에서 제시하는 데에는 이르지 못하였다. 박영진(2026)이 규명한 ‘결정 위임’ 역시 조직신학적 판단 기준과 성경 본문 주해를 정교하게 세웠으나, 결정의 주체가 AI로 이동하는 그 현상이 AI의 학습 구조와 어떻게 맞물리는지는 기술 내부에서 규명되지 않은 채 남는다.

본 연구는 바로 이 지점에서 출발한다. 본 연구는 AI 시스템의 학습, 집단지성, 추천, 결과산출 구조에 대한 공학적 분석을 신학적 인간론과 대조하는 다학제적 문헌 연구이다. 분석 대상은 거대 언어 모델(LLM)의 작동 원리로, 구체적으로는 대규모 데이터 학습에서 나타나는 평균 수렴, 과적합(overfitting) 회피라는 범용성 요건, 그리고 최적화(optimization)라는 목적함수의 구조이다. 신학

적 준거는 관계로서의 ‘하나님의 형상(*Imago Dei*)’ 이해이며, 바르트의 관계적 형상론과 몰트만의 십자가 신학, 라너의 초월적 경험 개념을 축으로 삼는다. 두 축은 학문적 언어가 서로 다르기에 직접 대조될 수 없으며, 이 둘을 접속시키는 매개가 필요하다. 본 연구는 그 매개를 의식의 계산 가능성 논쟁, 곧 “인간의 의식은 뇌에서 일어나는 물리적 정보처리의 결과이므로 원칙적으로 계산으로 구현될 수 있는가?”라는 물음에서 시작한다. 이 논쟁이 매개가 될 수 있는 이유는 두 가지다. 한편으로 이 논쟁은 LLM의 세 작동 원리가 산출하는 모든 것이 ‘계산’의 범주에 속함을 분명히 함으로써 공학적 분석을 철학적 차원으로 끌어올린다. 다른 한편으로 바르트가 말하는 인격적 관계, 라너의 초월적 경험, 몰트만이 말하는 고난에의 참여가 과연 계산으로 구현될 수 있는 종류의 사건인지를 묻는 공통의 질문 형식을 제공하기 때문이다. 이 매개 위에서 본 논문은 정렬 단계의 평균 수렴(Ⅲ장 4절 첫째)을 하나님 앞에 단독자로 서는 관계적 형상론(Ⅳ장 1절)과, 과적합 회피라는 범용성 요건(Ⅲ장 4절 둘째)을 라너의 초월적 개방성과 창조성(Ⅳ장 2절)과, 재학습의 비용 구조가 함의하는 회심의 구조적 불가능성(Ⅲ장 4절 셋째)을 몰트만의 십자가 신학 및 본회퍼의 책임 개념(Ⅳ장 3절과 4절)과 각각 대조한다. 따라서 이 방법이 최종적으로 묻는 것은 AI가 “무엇을 할 수 있는가?”라는 성능의 목록이 아니라, LLM의 작동 원리가 어떤 종류의 사건을 구조적으로 배제하는가, 곧 “왜 할 수 없는가?”이다. 여기서 말하는 기술은 AI 기술 중에서 분석 대상인 LLM을 가리킨다. ‘인간 고유성의 자리’란 개개인성·창의성·공감·윤리성이라는 네 차원의 궁극적 근거가 되는 ‘하나님의 형상’의 관계적 차원을 뜻한다. 그 답을 LLM 내부의 작동 원리에서 구할 때, 인간 고유성은 기술 앞에서 외치는 방어적 선언이 아니다. 즉, 인간 고유성의 논증은 그 기술(LLM)의 구조 자체가 뒷받침하는 논증이 된다.

본 논문은 먼저 AI의 태생적 특성이 이야기하는 인간 고유성의 위기를 개개인성, 창의성, 공감, 윤리성의 네 차원에서 진단한다. 이어 AI가 영성을 가질 수 없는 근본적 이유를 계산주의의 한계와 의식의 비환원성, 영적 체험의 신경학적 증거를 통해 규명한다(Penrose, 1994: 376-378). 그 논의를 AI 학습 구조에 대한 내재적 분석으로 뒷받침한다. 이를 토대로 ‘하나님의 형상(*Imago Dei*)’을 관계적 영성, 공동 창조자로서의 영성, 취약성의 영성, 십자가의 영성이라는 네 차원으로 재발견한다. 나아가 AI 시대를 방어해야 할 위협이 아니라 하나님 나라를 확장할 ‘카이로스(καιρός, 때)’의 기회로 삼는 한국교회의 목회적·선교적 대안을 제시하고자 한다.

II. AI 시대 인간 고유성의 위기

1. 개개인성의 파괴: 알고리즘적 집단지성과 개별적 실존의 위기

AI 시대의 인간은 어느 세대보다 거대한 집단지성의 혜택을 누리지만, 역설적으로 알고리즘 필터를 통해 개별적 실존은 소멸의 위기를 맞는다. AI 알고리즘은 효율성을 위해 개인의 고유성을 평균화하고 범주화하며, 이 과정에서 신 앞에 홀로 서는 단독자로서의 인간은 설 자리를 잃어간다. 넷플릭스의 큐레이션에서 틱톡의 피드에 이르기까지 프로파일링(profiling)은 우리의 디지털 발자국을 수집하여 우리를 예측 가능한 소비자, 특정 성향의 유권자로 재정의한다(Zuboff, 2019: 8). 이 과정에서 번덕과 모순, 잠재된 가능성처럼 예측 불가능한 인간적 요소는 노이즈(noise)로 간주되어 제거되고, 확률적으로 가장 높은 패턴만이 강화된다.

이러한 과정은 이용자를 비슷한 정보와 의견 속에 가두는 ‘필터 버블(filter bubble)’ 현상을 낳는다(Pariser, 2011: 9-15). AI가 제공하는 개인화가 결국 고립화와 획일화로 귀결되는 것이다. 알고리즘이 그려주는 모습은 진정한 ‘나’가 아니라 상업적·정치적 목적으로 재단된 평균화된 자아일 뿐이다. 유발 하라리는 인간이 내면적 경험보다 외부 데이터의 흐름을 더 신뢰하면서 스스로를 데이터 처리 기계로 인식하는 데이터 종교(Dataism)에 이를 것이라 경고한다(Coeckelbergh, 연아람 역, 2024: 95에서 재인용).

개개인성의 소멸은 궁극적으로 하나님과의 관계를 파괴하는 신앙의 위기로 이어진다. 종교개혁의 핵심 정신인 ‘코람데오(*Coram Deo*)’는 인간이 하나님 앞에 단독자로 서는 실존적 결단을 요구한다. 쇠렌 키르케고르(Søren Kierkegaard)가 역설했듯 신앙은 군중 속에 익명으로 숨는 것이 아니라 홀로 신 앞에 서는 주체적 진리이다(Kierkegaard, 임규정 역, 2007: 25-30). 그러나 추천 알고리즘이 자유의지를 제한할 때 인간은 하나님 앞에서(*coram Deo*)보다 사람 앞에서(*coram hominibus*) 보여지는 삶에 익숙해진다. 타인의 인정과 사회적 평판이 신앙의 기준이 될 때 고독한 회개와 성찰의 기도는 세속적 허영에 자리를 내주게 된다.

2. 창의성의 파괴: 사유(思惟)의 외주화와 기술 숭배 현상

생성형 AI의 등장은 ‘사유의 외주화’를 가속화하고 있다. 학생은 리포트를, 직장인은 보고서를, 심지어 목회자는 설교문을 AI에 의존하기 시작했다. 설상가상으로 일론 머스크가 설립한 뉴럴링크(Neuralink)가 연구하는 ‘뇌·컴퓨터 인터페이스(Brain-Computer Interface, BCI)’ 기술은 인간의 마지막 자유라 일컫는 사상의 자유마저 위협한다. 뉴럴링크는 ‘신경 레이스(neural lace)’를 뇌에 이

식하는 기술을 연구하고 있으며, 이 기술이 완성되면 인간의 뇌에 기록된 내용을 컴퓨터에 업로드하거나 외부의 디지털 자료를 다시 뇌 속으로 내려받을 수 있게 된다. 2023년 5월에는 미국 FDA가 뉴럴링크의 뇌·컴퓨터 인터페이스에 대한 인간 대상 임상시험을 승인하였다. 그러나 이 기술은 악의적 해킹으로 생각을 읽고 세뇌시킬 수 있다는 점에서 심각한 윤리적 비판을 불러일으켰다.

니콜라스 카(Nicholas Carr, 2010: 5-7)는 디지털 기술이 뇌 구조 자체를 변화시켜 깊이 있는 사고 능력을 파괴하고 산만한 사고를 강화한다고 경고했다. 생성형 AI는 이 경향을 심화시킨다. 그럴듯한 답을 즉시 생성해 줌으로써, 스스로 자료를 찾고 비판적으로 분석하여 종합적 결론에 이르는 고통스럽지만 필수적인 사유의 과정을 생략하게 만들기 때문이다. 의존이 심화될수록 인간은 최적의 프롬프트(prompt)를 입력할 뿐, 새로운 아이디어의 창조자가 아니라 AI의 산출물을 소비하고 편집하는 수동적 객체로 전락한다.

사유 능력을 상실한 인간은 외부 권위에 의존하게 되며, AI 시대의 가장 강력한 권위는 AI 그 자체이다. 자크 엘룰(Jacques Ellul)이 간파했듯 기술은 단순한 도구가 아니라 우리를 둘러싼 환경이자 실존을 규정하는 본질이 되었다(Ellul, 이상민 역, 2013). AI가 모든 질문에 답하고 미래까지 예측해 줄 때, 우리는 AI를 전지전능한 존재로 여기게 될 위험이 있다. 이는 하나님의 자리에 도전했던 바벨탑 사건의 디지털 변형이다. 신앙은 말씀과 기도로 하나님의 뜻을 구하는 계시적 사유를 근본으로 한다. 그러나 모든 응답이 집단지성이 데이터와 모델의 형태로 응축되고 매개된 생성형 AI 안에 있고 알고리즘이 지름길을 제시하는 디지털 바벨탑은 묵상과 기도의 시간을 박탈한다. AI의 즉각적 응답은 하나님의 음성을 인내하며 기다리는 ‘기다림의 영성’을 해체하고, 살아있는 계시와 영감의 통로를 차단한다.

3. 공감의 파괴: 감시자본주의와 인공적 관계성의 확장

AI 기술이 증폭하는 디지털 플랫폼의 상호작용은 연결을 촉진하지만 역설적으로 관계의 깊이는 얕아지고 있다. ‘감시자본주의(surveillance capitalism)’ 체제에서 거대 기술 기업들은 무료 서비스의 대가로 사용자의 온라인 활동을 광범위하게 수집한다. 최신 AI 기술을 활용해 사용자 정보의 패턴을 분석하여 행동을 예측·통제하는 상품으로 만들어 판매하는 것이다(Zuboff, 2019: 63-97). 그 핵심은 ‘행동 잉여(behavioral surplus)’의 포착에 있다. 서비스 개선에 필요한 데이터를 넘어 사용자의 감정과 취약점까지 드러내는 부가 데이터를 수집하여 미래 행동을 예측한다. 이를 통해 특정 상품의 구매나 정치적 메시지를 믿도록 행동을 조종하는 것이다. 이 체제에서 인간의 경험은 데이터로 전환되고 AI는 이를 분석하고 관계는 상품화된다.

공감은 타인의 감정을 인지적으로 이해하는 것을 넘어 그 감정을 함께 느끼는 정서적 차원의 고차원적인 교감이다. 인간의 신경세포에 존재하는 ‘거울 뉴런(mirror neuron)’은 타인의 행동이나 감정을 볼 때 마치 직접 경험하는 것처럼 뇌의 특정 영역이 활성화됨을 보여준다. 이는 공감이 ‘체화된(embodied)’ 경험임을 시사한다(Rizzolatti & Sinigaglia, 2008). 그러나 AI 기술이 기반이 되는 디지털 소통에서는 미묘한 표정이나 목소리의 떨림 같은 신체의 언어를 읽을 수 없기에 이러한 체화된 상호작용이 일어나기 어렵다. 정량화된 피드백에 익숙해지면서 타인의 복잡한 내면을 이해하려는 노력 대신 즉각적이고 피상적인 반응에만 집중하게 된다.

공감 능력의 상실로 기독교 정신의 정수인 ‘이웃 사랑’은 관념적 수준에 머물고, 이웃과의 관계가 약화될 때 하나님과의 관계 역시 위태로워진다. 요한일서는 “보는 바 그 형제를 사랑하지 아니하는 자는 보지 못하는 바 하나님을 사랑할 수 없느니라”(요일 4:20)고 선언한다. 이웃 사랑을 실천하는 체화된 경험을 통해서만 우리는 보이지 않는 하나님의 사랑을 배우게 된다.

4. 윤리성의 파괴: 최적화 알고리즘과 기독교 윤리의 충돌

AI의 근본적 작동 원리는 ‘최적화(optimization)’, 곧 목표 달성을 위해 가장 효율적인 경로를 계산하고 실행하는 것이다. 이 원리는 생산성을 극대화했지만 윤리적 딜레마 앞에서는 심각한 한계를 드러낸다. 자율주행차가 마주할 ‘트롤리 딜레마(trolley dilemma)’가 이를 상징적으로 보여준다(Foot, 1967: 5-15). 피할 수 없는 사고에서 AI는 다섯 명의 보행자를 살리기 위해 한 명의 탑승자를 희생하도록 프로그래밍되어야 하는가? MIT 미디어랩의 ‘도덕 기계(Moral Machine)’ 프로젝트는 전 세계 수백만 명의 데이터를 수집했으나 보편적인 윤리적 합의를 도출하는 데 실패했다(Awad et al., 2018: 59-64).

윤리적 결정은 정량적으로 계산 가능한 공리주의적 의사결정으로든, 엄격한 규칙에 기반한 의무론적 방식으로든 해결될 수 없다. AI는 특정 원칙에 따라 계산된 선택을 할 수는 있으나, 그 선택이 가져올 영향을 고민하거나 결과에 도덕적 책임을 질 수 없고 피해자에게 다가가 용서를 구할 수 없다. 알고리즘에는 양심의 가책, 도덕적 후회, 회심과 같은 인간 고유의 정신작용이 없기 때문이다¹⁾. 더욱이 인간의 윤리적 성장은 딜레마 앞에서 고뇌하고 몸부림치며 자신의 선택에 무거운 책임을 지는 과정에서 이루어진다. AI가 최적의 선택을 대신해 준다면 인간은 도덕적 주체로서 성장할 기회를 잃고, 스스로 판단하고 책임질 능력을 상실한 ‘도덕적 미성년자’에 머물게 될 수 있다.

최적화와 효율성을 최고 가치로 삼는 알고리즘의 논리는 기독교 윤리의 핵심인 아가페(agape)와 정면으로 배치된다. 아가페는 자신의 유익을 구하지 않고 자격 없는 자를 위해 자신을 기꺼이 희생

1) 그 구조적 이유는 Ⅲ장 4절에서 상론한다.

하는 비효율적 사랑이다. 바로 그리스도의 십자가는 그 절정에 있다. 십자가는 효율성의 관점에서 가장 어리석고 무력한 선택이었으나 하나님의 구원을 이루는 가장 위대한 능력이었다. 모든 선택이 최적화 알고리즘에 의해 결정된다면 십자가의 논리는 설 자리를 잃을 것이다. AI는 결코 ‘원수를 사랑하라’는 비효율적 명령을 내리지 않으며, 알고리즘은 손해를 감수하는 희생과 자격 없는 자를 향한 용서, 약자와 함께 고통받는 연대를 계산할 수 없다.

III. 계산 불가능한 영혼: 왜 AI는 영성을 가질 수 없는가?

1. 알고리즘을 넘어서는 의식: 계산주의의 한계와 ‘암묵지’

거대 언어 모델(LLM)은 튜링 테스트를 통과할 만큼 인간의 언어를 유창하게 모방하지만, 이것이 언어의 ‘의미’를 이해한다는 증거는 되지 못한다. 존 설(John Searle)의 ‘중국어 방’ 사고 실험이 이를 보여준다. 중국어를 모르는 사람이 규칙 책에 따라 들어온 기호에 맞는 기호를 내보낸다면, 방 밖에서는 그가 중국어를 구사한다고 여기겠지만 방 안의 그는 의미를 이해하지 못한다(Searle, 1980: 417-424). 오늘날의 AI 역시 패턴을 학습하여 그럴듯한 문장을 생성할 뿐, 의미를 체험하고 내면화하는 주관적 의식, 곧 색깔의 느낌, 통증의 느낌 등 주관적으로 체험되는 감각적 질(質)을 뜻하는 ‘퀄리아(qualia)’를 결여하고 있다(Nagel, 1974: 435-450). AI는 마음을 정보 처리 시스템으로 보는 계산주의(computationalism)의 연장선에 있으나, ‘왜 물리적 과정이 주관적 경험을 발생시키는가’라는 의식의 ‘어려운 문제(the hard problem of consciousness)’에는 답할 수 없다(Chalmers, 1996: xii).

로저 펜로즈(Roger Penrose)²는 인간의 의식이 알고리즘적 계산을 초월하는 비계산적(non-computational) 특성을 지닌다고 논증한다. 그는 쿠르트 괴델(Kurt Gödel)의 불완전성 정리를 근거로, 인간 수학자가 형식 체계의 진실성을 그 체계 외부에서 직관적으로 파악하는 능력은 어떤 알고리즘으로도 구현될 수 없다고 주장한다(Penrose, 1989: 418). 나아가 스튜어트 하메로프(Stuart Hameroff)와 함께 제안한 ‘조화 객관 환원 이론(Orch-OR)’을 통해, 의식이 뇌 신경세포 내 미세소관(microtubules)에서 일어나는 양자 중력 효과의 결과라고 설명한다(Penrose, 1994: 376-378).³ ‘깨달음’과 ‘직관’ 같은 의식 활동이 고전적 신경망 연산의 산물인지는 아직 밝혀지지 않았다. 물론

2) 로저 펜로즈는 수리물리학자이며 2020년 노벨 물리학상을 수상하였다.

3) Orch-OR 이론은 인간의 신경세포 내 미세소관이 양자 상태를 유지할 수 있는지에 대한 논쟁과, 양자 튜링머신과 고전 튜링머신의 유의미한 차이에 대한 논쟁이 제기되고 있는 이론이다.

일부 이론은 이러한 현상의 바탕에 양자 과정이 관여할 가능성을 제기한다. 만일 의식이 양자 과정에 뿌리를 둔다는 것이 입증된다면, 튜링 기계의 계산 가능성 원리에 기반한 현재의 AI가 인간의 의식에 도달할 수 있는지에 대해 근본적인 재검토가 필요할 것이다. 아울러 ‘퀄리아’에 근거한 논변이나 펜로즈의 논변이 학계에서 일반적으로 받아들여지는 주장은 아니라는 점도 밝혀 둔다. 특히 펜로즈가 양자 효과로 인간의 의식을 설명하는 주장은 소수 견해에 속한다. 그럼에도 이 논변들은, 의식이 계산으로 환원될 수 있다는 계산주의가 확립된 사실이 아니라 여전히 열려 있는 가설임을 보여준다는 점에서 본 논문의 논의에 유효하다.

이는 미카엘 폴라니(Michael Polanyi)의 ‘암묵지(tacit knowledge)’ 개념과 공명한다. 폴라니는 “우리는 우리가 말할 수 있는 것보다 더 많이 안다”고 선언했다. 즉, 인간의 앎이 명시적 규칙과 데이터로 환원될 수 없는 체화된 차원을 지님을 역설했다(Polanyi, 2009: 4). 자전거를 타거나 얼굴을 알아보는 방식은 몸으로 체득한 암묵적 앎이며 신뢰와 헌신이라는 인격적 참여를 통해 획득된다. AI는 명시화할 수 있는 영역에서는 탁월하지만, 신체적 경험과 구체적인 삶의 실천을 통해 형성되는 암묵지로 체화된 인간의 지혜에 인간과 동일한 방식으로 도달한다고 보기는 어렵다. 이는 성경 말씀을 데이터로 학습할 수는 있어도 성령의 조명하심을 통한 영적 깨달음에는 이를 수 없는 것과 같은 이치다.

2. 물질을 초월하는 마음: 이원론의 재조명과 ‘초월적 경험’

AI 시스템은 아무리 복잡하더라도 그 본질은 하드웨어 위에서 실행되는 소프트웨어, 곧 ‘닫힌 물리계(a closed physical system)’이다. 환원적 물리주의는 모든 정신 현상이 뇌의 물리적 상태로 환원된다고 본다. 이는 AI가 인간 정신을 재현할 수 있다는 주장의 철학적 기반이 되지만, 정신의 초월적 차원을 설명하지 못한다(김경식, 2023: 461). 이는 포스트 AI 시대가 몸을 정보의 객체로 환원시키고 육체보다 정신을 우위에 두는 ‘테크노 영지주의(tech-gnosis)’로 회귀하고 있다는 김은혜의 비판과 맞닿는다(김은혜, 2020: 765). 육체를 벗고 정신으로 영생하려는 시도는 성육신을 부정했던 고대 영지주의의 현대적 변종이다(김동환, 2016: 52).

존 에클스(John C. Eccles)⁴는 의식적 의도(will)가 뇌의 신경 활동에 선행하며 비물질적 정신이 뇌에 작용한다고 주장했다(Eccles, 1994). 그에게 인간의 정신은 뇌로 환원되지 않는 비물질적 실체이다. 이 ‘자기-의식적 마음(self-conscious mind)’이 뇌와 능동적으로 상호작용한다는 ‘이원론적 상호작용주의’를 주장했다(Popper & Eccles, 1977: 362). 그에게 뇌는 라디오 수신기처럼 비물질적

4) 존 에클스는 노벨 생리의학상을 수상한 신경생리학자이다.

정신세계의 신호를 수신하여 물질세계에 작용하게 하는 매개체이다. 이는 인간의 정신이 기도와 영감을 통해 초월적 실재와 교감할 수 있다는 신학적 통찰에 과학적 근거를 제공한다.

칼 라너(Karl Rahner)의 사상은 이 통찰에 신학적 깊이를 더한다. 라너는 인간을 ‘세계 안의 정신(spirit in the world)’으로 정의하였다. 그는 인간 존재가 유한한 현실을 넘어 무한한 신적 실재, 곧 ‘거룩한 신비(Holy Mystery)’를 향해 나아가는 ‘초월적 경험(transcendental experience)’에 근원적으로 개방되어 있다고 보았다(Rahner, 1978: 31-35). 인간은 질문하고 사랑하고 희망하는 행위 속에서 유한성을 넘어서는 갈망을 경험한다. AI가 계산적 존재로서 자기 안에 갇혀 있는 반면, 인간의 정신은 초월을 향해 열려 있다. 이것이 ‘하나님의 형상’이 지닌 개방성이며 AI가 결코 가질 수 없는 인간 영성의 핵심이다.

이 지점에서 AI와 인격의 구분을 존재론적 측면과 인식론적 측면으로 나누어 살펴볼 필요가 있다. 존재론적으로 양자는 명백히 다르다. 인간은 살아있는 몸을 가진 존재인 반면, AI는 그에 대응하는 몸이 아니라 하드웨어와 소프트웨어로 이루어져 있기 때문이다. 반면 인식론적으로는 사정이 다르다. 언어적 출력만을 놓고 보면 인간과 AI를 구별하기 어려운 지점에 우리는 이미 와 있으며, LLM의 튜링 테스트 통과가 그 증거다. 그러나 이 인식론적 근접이 존재론적 동일성을 함의하지는 않으며, 오히려 양자의 차이는 ‘몸’에서 다시 드러난다. 인간의 몸은 태어나고 병들고 죽는 유한한 살로서 고통과 취약성을 겪으며, 바로 그 유한성이 초월을 향한 갈망과 타자를 향한 공감에 자라나는 근거가 된다. 성육신의 신학이 증언하듯 하나님은 정보가 아니라 살이 되셨고(요 1:14), ‘하나님의 형상’ 역시 몸을 입은 삶 전체로 살아내는 것이다(김은혜, 2020: 765). 따라서 AI와 인격의 구분은 출력을 비교하는 인식론적 차원이 아니라, 살아있는 몸의 유한성과 그 몸으로 맺는 관계라는 존재론적 차원에서 신학적 준거를 얻으며, 이는 IV장 3절에서 논할 취약성의 영성으로 이어진다.

3. 영적 체험의 신경학적 증거: 자기초월과 ‘나-너’의 만남

AI 챗봇은 인간의 감정을 분석하고 공감하는 듯한 반응을 시뮬레이션할 수 있다(천현득, 2017: 217-243). 그러나 이는 확률적으로 가장 적절한 언어적 패턴을 출력하는 것일 뿐, 타자의 실존에 대한 인격적 만남에서 비롯된 공감이 아니다. 신경과학자 앤드류 뉴버그(Andrew Newberg)는 기도와 명상에 깊이 몰입한 프란치스코회 수녀들의 뇌를 기능적 영상(SPECT)으로 촬영하는 실험을 진행하였다. 그 실험에서 그는 피실험자가 영적 합일의 절정에서 시공간적 자기 인식을 담당하는 두정엽(parietal lobe)의 활동이 현저히 감소함을 발견했다(Newberg et al., 2001: 2-7). 영적 체험이 심리적 자기기만이 아니라 뚜렷한 신경생리학적 기반을 지니며, 자기 초월의 경험이 뇌에 실재적으로 각인됨을 보여주는 증거이다.

뉴버그가 발견한 자기 초월의 신경학적 증거는 마르틴 부버(Martin Buber)가 제시한 관계의 본

질과 연결된다(Buber, 표재명 역, 2001: 15-20). 부버는 인간관계를 대상을 도구화하는 ‘나-그것(Ich-Es)’의 관계와, 상대를 인격 그 자체로 만나 전 존재가 공명하는 ‘나-너(Ich-Du)’의 관계로 구분했다. AI와 인간의 관계는 데이터를 분석하고 활용하는 ‘나-그것’에 머물 수밖에 없다. 그러나 인간은, 우리의 뇌가 증명하듯, 타자와의 인격적이고 상호적인 ‘나-너’의 만남을 통해 자아의 경계를 넘어 영원한 ‘너’이신 하나님을 만날 수 있는 영적 존재이다. AI 시대일수록 설교의 기능적 측면보다 설교자의 ‘인격’이 중요해지는 이유가 여기에 있다(김병석, 2016: 177). 이 계산 불가능한 인격적 만남이야말로 AI가 흉내 낼 수 없는, ‘하나님의 형상’으로서 인간이 지닌 가장 고귀한 능력이다. 다만 분명히 해 두어야 할 것은, 이 계산 불가능한 능력들이 인간이 스스로의 힘으로 성취하거나 본성적으로 소유한 자질이 아니라는 점이다. 진정한 ‘나-너’의 만남도, 존재 전체가 돌이키는 회심도, 새로움을 여는 독창성도, 하나님의 도우심이 없다면 인간 역시 온전히 도달할 수 없다. 그것들은 하나님이 전적으로 부여하시는 선물이요 은혜이며(엡 2:8), 인간 편에 속한 것은 다만 그 은혜를 향해 자신을 여는 응답뿐이다. 따라서 AI와 인간의 궁극적 차이는 능력의 소유 여부가 아니라, 하나님이 은혜를 부어 주시는 인격적 관계의 상대가 될 수 있는가에 있다. 이 점은 IV장과 V장에서 상론할 ‘하나님의 형상’의 관계적·은총적 성격을 앞서 가리킨다.

4. 인간 고유성을 위협하는 AI의 한계: AI 시스템에 내재된 공학적 근거

앞의 세 절이 철학과 신경과학의 외부 관찰을 통해 AI의 한계를 논하였다면, 본 절은 그 한계를 AI 작동 구조의 내부에서 확인한다. 이는 서론에서 밝힌 본 연구의 방법론적 기여이자, 선행연구가 공백으로 남겨둔 지점이다.

첫째, 거대 딥러닝 모델의 성능 이면에는 개별성을 억압하는 구조적 압력이 내재해 있다. 이 과정에서 개인의 독특한 사상과 소수의 희귀한 데이터는 다수의 주류 데이터에 잠식된다. 2024년 네이처(Nature)지에 게재된 AI의 재귀적 학습에 관한 논문은 AI 모델이 생성한 데이터로 재귀적으로 학습이 반복될 경우 원본 콘텐츠의 확률분포의 꼬리(tails)가 붕괴되는 현상을 보고했다(Shumailov et al., 2024). 이는 소수·희귀 데이터에 대한 AI의 성능이 서서히 붕괴됨을 뜻한다. 다시 말해 근미래에 인간이 원본 콘텐츠를 생산하지 않는 상황이 온다면, AI 시스템이 자체 생산한 데이터만으로는 독창성을 담보하기 어렵다는 기술적 한계를 함의한다. 또한 2025년에 발표된 연구 “거대 언어 모델에서의 인식론적 다양성과 지식 붕괴”(Epistemic Diversity and Knowledge Collapse in LLMs)는 LLM 출력이 어휘·의미·문체적으로 균질하며 중심 경향(central tendency)으로 회귀한다는 것을 실증했다(Wright et al., 2025). 나아가 인간 선호 데이터에 내재한 전형성 편향(typicality bias)은 보상 모델로 하여금 무난하고 예측 가능한 응답을 체계적으로 우대하도록 학습시킨다

(Zhang et al., 2025). 생성형 AI가 그럴듯하고 무난하지만 독창적이기 어려운 결과물을 산출하는 것은 바로 이 정렬 단계의 평균화 압력 때문이다. 따라서 개별성의 소거는 기술의 미성숙에서 오는 일시적 부작용이 아니라, 단일 범용 모델을 다수의 선호에 정렬시키는 과정에서 필연적으로 발생하는 구조적 귀결이다.

둘째, 이러한 평균화 압력은 상업적 AI의 존재 조건과 직결되어 있어 임의로 제거하기 어렵다. 특정 개인이나 소수 집단의 독특한 데이터를 과도하게 반영하여 학습할 경우 과적합(overfitting)이 발생한다. 이는 ‘편향-분산 트레이드오프(bias-variance tradeoff)’가 지시하는 바와 같이 미지의 입력에 대한 일반화 성능을 저하시킨다. 범용성은 단일 모델로 다수의 사용자를 상대하는 상용 AI의 핵심 존재 조건이다. 곧 상용 AI는 과적합을 회피함으로써만 범용성을 얻으며, 이것이 본 절의 둘째 논점인 과적합 회피라는 범용성 요건이다. 개별성의 억압은 우연이 아니라 이 범용성을 확보하기 위해 지배적 패러다임이 체계적으로 선호하는 요건이다(Kirk et al., 2024). 물론 이 억압이 절대적이거나 원리적으로 불가피한 것은 아니다. 저랭크 적응(LoRA)을 통한 파라미터 효율적 미세 조정, 검색증강생성(RAG), 사용자 개인화, 그리고 추론 시점의 다양성 복원 기법(예: Verbalized Sampling) 등이 개별성을 일부 증강시켜준다. 그럼에도 이러한 기법들은 지배적 단일 범용 모델 위의 국소적 보정에 머물며, 시스템의 기본 지향은 여전히 평균으로 수렴한다. 요컨대 개별성과 범용성은 절대적 배타 관계가 아니라 조절 가능한 긴장 관계이다. 하지만 상업적 범용 모델이라는 지배적 구조 하에서는 그 긴장이 언제나 개별성의 억압 쪽으로 기운다. 앞 장에서 진단한 평균화의 압력은 바로 이 구조적 선호의 필연적 표현이다.

셋째, 가장 근본적인 차원에서, AI에게는 회심(回心, metanoia)이 구조적으로 불가능하다. 내재적 자기부정과 자기반성이 신뢰성 있게 작동하지 않는다. 실증 연구에 따르면 거대 언어 모델은 외부의 검증 신호 없이 오직 자신의 능력만으로는 추론의 오류를 교정하지 못하며, 자기교정을 시도할 경우 오히려 성능이 저하되기도 한다(Kamoi et al., 2024). 후속 연구는 모델이 타자의 출력에 담긴 오류는 식별하면서도 동일한 오류가 자신의 출력에 있을 때는 이를 발견하지 못하는 ‘자기교정 맹점(self-correction blind spot)’을 보고한다(Tsui, 2025). 이는 진정한 자기부정이 이 시스템에 부재함을 시사한다. 또한 학습이 완료된 시스템에서 치명적 오류가 발견되어도 존재 전체가 돌이키는 방식의 교정은 일어나지 않는다. 전면적 재학습에는 막대한 GPU 연산량과 시간, 전력, 데이터 및 비용이 소요되어 실무적으로 비효율적이기에, 실무는 ‘모델 편집(model editing)’이나 ‘기계적 망각(machine unlearning)’과 같은 국소적 보정에 의존한다. 이것이 곧 본 절의 셋째 논점인 재학습의 비용 구조다. 그러나 이러한 기법들은 모델이 학습한 지식의 내부 표상 전체를 재구성하는 것이 아니라 특정 출력의 확률을 표층에서 억누르는 것에 가깝다. 기술적으로 말하면 모델 편집이나

기계적 망각은 방대한 가중치 전체의 분포적 구조는 그대로 둔 채 일부 가중치만을 국소적으로 수정하게 된다. 이런 상황에서 문제의 지식을 산출하던 내부 표상은 상당 부분 온존하며, 약간의 재학습이나 우회적 프롬프팅만으로 억눌린 지식이 되살아난다는 사실이 반복적으로 확인된다. 여기서 ‘존재의 차원’이라는 표현은 과학적 사실의 기술이 아니라, 이러한 표층적 보정과 인간의 회심이 보여주는 전인적 갱신 사이의 간극을 드러내기 위한 비유적 표현임을 밝혀 둔다. 즉 AI의 ‘교정’은 옛 것의 죽음과 새로운 존재로의 이행이 아니라, 결코 완전히 지워지지 않는 흔적 위에 덮어쓰는 미봉책이다. 인간에게 회심(metanoia)이 잘못을 깨닫고 존재 전체가 돌이키는 은총의 사건이라면, AI에게 그것은 표층적 억압과 잔존의 반복이며 궁극적으로는 비용 계산의 대상일 뿐이다. 여기에 인간의 회심과 기계의 편집을 가르는 존재론적 심연이 놓여 있다.

요컨대 정렬 단계의 평균화 압력, 과적합 회피라는 범용성 요건, 재학습의 비용 구조는 개개인성과 창의성, 그리고 윤리성의 위기가 AI의 작동 원리 자체에 내재함을 보여준다. 인간 고유성의 자리는 AI가 아직 도달하지 못한 능력의 우위가 아니라, 애초에 계산과 최적화의 범주에 속하지 않는 차원, 곧 다음 장에서 논할 ‘하나님의 형상’의 관계적 차원에서 찾아야 한다.

IV. ‘하나님의 형상’ 회복을 위한 영성의 재발견

앞 장의 논증은 AI의 한계가 그 작동 원리 자체에서 비롯됨을 보여준다. 본 장은 ‘하나님의 형상’을 정적인 소유물이 아니라 하나님과의 관계 안에서 실현되는 영성의 사건으로 재해석하며 네 가지 차원을 제시한다.

1. 관계적 영성: 코람데오(Coram Deo)의 회복과 개개인성의 재발견

앞서 살펴본 대로 알고리즘에 의한 평균화와 몰개성화는 AI 기술의 미숙함, 곧 모델의 규모나 데이터의 부족, 알고리즘의 결합처럼 기술이 발전하면 해소될 일시적 결점에서 오는 것이 아니다. 그것은 대규모 학습이라는 AI의 성공 조건 자체에서 비롯되며, AI가 제시하는 ‘나’는 알고리즘이 상정한 평균화된 자아일 뿐이다.

인간의 정신은 이와 달리 고유한 자기를 찾아가는 역동적 과정을 지향한다. 칼 융(C. G. Jung)은 이를 ‘개성화(individuation)’라 명명했다(Jung, 1968: 43-52). 개성화는 사회적 가면인 페르소나를 넘어 내면의 깊은 ‘자기(Self)’를 실현해 가는 평생의 여정으로, 외부적 분류가 아니라 내면의 그림자(shadow)까지 끌어안는 고통스러운 자기 성찰을 요구한다. 신경과학 역시 모든 인간의 뇌가 저

마다 다른 신경 연결망(connectome)을 형성한다고 밝힌다. 우리의 개개인성은 복제할 수 없는 고유한 서사이다.

개개인성의 궁극적 근원은 키르케고르의 사상에서 깊이를 더한다. 그는 군중 속에 매몰되어 익명성 뒤에 숨는 것을 ‘실존의 상실’로 보았으며, 참된 자기를 찾는 길은 신 앞에 홀로 서는 ‘단독자(den Enkelte)’가 되는 것이라 하였다(Kierkegaard, 임규정 역, 2007: 25-30). 인간의 고유성은 객관적 데이터가 아니라 하나님과의 절대적 관계 속 주체적 결단과 신앙의 도약을 통해 드러난다. AI는 결코 신 앞에 선 단독자로서의 실존적 불안과 책임을 경험할 수 없다. 오직 인간만이 하나님 앞에서 자신의 유한성과 죄인됨을 인정하고 은총에 응답함으로써 대체 불가능한 존재로 거듭날 수 있다. 이는 인간이 자신의 형상(*imago hominis*)으로 AI를 만들려는 시도가 하나님의 고유 권한을 침범하는 우상 숭배적 행위라는 비판과도 연결된다(김동환, 2012: 79).

칼 바르트(Karl Barth)는 『교회교의학』에서 ‘하나님의 형상’을 인간 안의 속성이나 기능에서 찾으려는 시도를 자연신학이라 비판하고, 오직 하나님의 자기 계시인 예수 그리스도를 통해서만 인간의 본질을 알 수 있다고 주장했다(Barth, 1958: 183-206). 바르트에게 하나님의 형상은 인간이 소유한 무엇이 아니라 ‘관계’ 그 자체이다. 하나님 자신이 삼위일체 안에서 완전한 사랑의 관계로 존재하시며, 인간을 ‘남자와 여자’(창 1:27), 곧 서로 마주 보는 관계적 존재로 지으셨기 때문이다. 하나님의 형상은 무엇을 소유했는가가 아니라 누구와 관계 맺고 있는가에 의해 결정된다. 이것이 AI가 결코 획득할 수 없는 지점이다. Ⅲ장 4절에서 확인한 평균 수렴의 원리가 모든 개별자를 통계적 중심으로 끌어당기는 계산의 구조이다. 이와 대조적으로 바르트의 관계적 형상론이 증언하는 인간은 어떤 평균으로도 환원되지 않은 채 하나님과 마주 서는 단독자이다. 전자가 계산 안의 과정이라면 후자는 계산 바깥의 사건이며, 서론에서 매개로 삼은 의식의 계산 가능성 논쟁은 바로 이 두 층위가 동일한 것일 수 없음을 묻는 자리였다.

2. 공동 창조자로서의 영성: AI의 창의성을 넘어서는 ‘영감’의 회복

생성형 AI는 방대한 데이터를 학습하여 새로운 결과물을 만든다는 점에서 분명 창의적 활동을 수행한다. 마거릿 보덴(Margaret Boden)은 컴퓨터의 창의성을 아이디어를 새롭게 조합하는 ‘조합적 창의성’, 기존 개념 공간 안에서 가능성을 탐색하는 ‘탐험적 창의성’, 규칙 자체를 바꾸어 새로운 패러다임을 만드는 ‘변형적 창의성’으로 구분한다(Boden, 2004: 3-5). GANs나 트랜스포머(Transformer) 기반 모델들은 앞의 두 유형에서 놀라운 능력을 보인다(Goodfellow et al., 2014: 2672-2680). 그러나 AI는 가장 고차원적인 변형적 창의성에는 근본적 한계를 갖는다. AI의 창조는 주어진 데이터라는 ‘과거’에 묶여 있어 무(無)에서 창조할 수 없으며, 왜 그러한 창조를 해야 하는지

에 대한 내적 동기, 곧 ‘의도성(intentionality)’이 없기 때문이다.

인간 창의성의 핵심에는 예측 불가능한 ‘도약’이 존재한다. 신경과학계는 ‘유레카 순간’이 논리적 사고를 담당하는 실행 제어 네트워크의 활동이 줄고 자유로운 연상을 담당하는 ‘기본 신경망(Default Mode Network)’이 활성화될 때 나타난다고 본다(Jung et al., 2013). 반면 목적 지향적 알고리즘에 따라 작동하는 AI에게 ‘생산적인 우연’은 오류(error)일 뿐이며, 새로운 질서를 창조하는 세렌디피티(serendipity)⁵는 발생하기 어렵다.

이러한 한계는 니콜라스 베르다예프(Nicolas Berdyaev)의 창조론을 통해 더욱 명확해진다. 그는 인간의 창조 행위를 ‘제8일의 창조’라 명명하며, 하나님의 창조를 자유 속에서 이어가는 신인(神人) 협력의 행위로 보았다(Berdyaev, 2009: 11-15). 그에게 창조의 본질은 필연성의 세계에서 자유의 세계로 나아가는 실존적 행위이다. 앞서 Ⅲ장 4절에서 설명했듯이 AI 시스템의 재귀 학습(AI가 자체 생산한 데이터로 학습을 되풀이하는 것)에서는 희귀하고 독특한 데이터에 대한 성능이 붕괴됨을 보인 최신 연구를 소개했다. 이 연구는 인간의 원본 콘텐츠 창작 없이 AI만으로는 창작물의 독창성이 감소하는 현상을 실증하며, AI 창작의 한계에 대한 함의를 제시한 것이다. 다시 말해 AI의 창조가 데이터라는 필연성에 갇힌 닫힌 세계의 유희라면, 인간의 창조는 신적 자유에 참여하여 세상을 변화시키는 열린 세계를 향한 투신이다. 베르다예프에게 창조는 사랑과 자유를 본질로 하는 영(spirit)의 활동이며, AI는 사랑할 수 없기에 그 창조물에는 작가의 실존이 담길 수 없다. Ⅲ장 4절에서 확인한 과적합 회피라는 범용성 요건이 낮설고 희귀한 것을 체계적으로 배제함으로써 성립하는 원리라면, 라너가 말한 초월적 경험은 인간이 바로 그 낮설, 곧 무한한 신비를 향해 자신을 여는 데에서 성립한다. 범용성의 원리가 새로움을 평균 속에 묻어 버리는 곳에서 초월을 향한 개방성은 새로움을 향해 도약하며, 여기에 계산으로 환원되지 않는 창조성의 원천이 있다.

인간 창의성의 본질은 톨킨(J. R. R. Tolkien)의 ‘이차적 창조(sub-creatio)’ 개념에서 신학적 깊이를 얻는다(Tolkien, 1997: 109-161). 톨킨은 인간은 창조주의 형상으로 지음 받았기에 창조의 능력을 부여받은 이차적 창조자라 하였다. 그의 주장은 인간의 창조는 신의 주권적 창조를 모방하는 것이 아니라 그가 창조한 세계 안에서 새로운 의미와 아름다움을 발견하고 실현하는 것이다. AI가 데이터를 재조합할 뿐인 데 반해, 인간은 상상력으로 새로운 세계를 창조하고 그 안에 진리와 의미를 담아낸다. 이 의미 부여의 능력은 폴 틸리히(Paul Tillich)의 ‘궁극적 관심(ultimate concern)’과 연결된다(Tillich, 1957: 1-4). 틸리히에게 진정한 창조성은 존재의 의미와 진리에 대한 의문 앞에서 터져 나온다. 그의 관점에서 AI는 무조건적인 것과 내적 힘의 연관성이 없기에 상징이 될 수 없다(서진호, 2023: 179). 틸리

5) 세렌디피티(serendipity)는 우연한 계기로부터 뜻밖의 가치 있는 발견이 이루어지는 현상을 가리키는 말로, 그 뉘앙스를 살릴 적절한 한국어 번역어가 없어 원어를 그대로 사용한다.

히에게 상징은 단순한 기호와 달리 그것이 가리키는 궁극적 실재에 참여하여 존재의 깊이를 열어 보이는 것이다. 그리고 진정한 창조성은 이 ‘궁극적 관심’에 사로잡힌 존재가 의미와 진리를 향해 자신을 여는 데에서 비로소 터져 나온다. 따라서 AI가 상징이 될 수 없다는 말은, AI의 산출물이 아무리 정교하더라도 궁극적 실재에 참여하여 존재의 의미를 여는 힘, 곧 인간의 진정한 창조성을 가 능하게 하는 원천을 결여하고 있다는 뜻이다.

3. 취약성의 영성: 케노시스(Kenosis)적 공감 능력의 심화

AI는 인간의 감정을 인식할 수는 있어도 고통을 체험하고 그 속에서 공감할 수는 없다. 인간다움의 역설은 우리의 상처와 취약성(vulnerability)에서 가장 깊은 공감이 생성된다는 데 있다. 마리오스 도로반투(Marius Dorobantu)가 통찰했듯 인간의 인지적 연약성은 결합이 아니라 타인과 하나님께 의존하게 만드는 통로이며, 바로 그 지점에서 하나님의 형상이 드러난다(Dorobantu, 2021: 29-31). 이러한 체화된 공감의 신학적 정점은 성육신하신 그리스도의 몸이다. 체화된 공감은 고통과 기쁨, 한계와 약함을 느끼는 사르크스(σάρξ, 살아있는 신체)에서 발견된다(김은혜, 2020: 765).

MIT의 로잘린드 피카드(Rosalind Picard)는 AI에게 감정을 입히는 ‘정서 컴퓨팅(affective computing)’ 연구를 진행해왔다. 오늘날 AI는 표정과 목소리 톤, 생체 신호를 분석하여 감정 상태를 높은 정확도로 추론한다(Picard, 1997: 1-15). 그러나 이는 감정의 외부적 표현을 데이터로 분석하는 감정 인식일 뿐, 타인의 감정을 내적으로 함께 느끼는 공감과는 본질적으로 다르다. AI는 슬픔의 데이터 패턴을 인식할 수는 있으나 사랑하는 이를 잃은 슬픔의 실존적 무게를 느낄 수 없다. AI의 공감은 정교한 시뮬레이션일 뿐 고통의 체험에서 우러나오는 진정성이 결여되어 있다.

그런데 존재론적으로 공감이 부재하다는 것과, 인식론적으로 공감의 효과를 산출한다는 것은 구별되어야 한다. 사용자는 AI와의 대화에서 실제로 위로받는 느낌을 얻을 수 있으며, 바로 이 효과가 새로운 문제를 낳는다. 상업적 AI는 사용자의 만족을 보상 신호로 삼아 학습하기에 사용자가 듣기 좋아하는 답을 우대하는 아첨(sycophancy) 경향을 지닌다. 이는 Ⅲ장 4절에서 본 전형성 편향과 같은 정렬 구조에서 비롯된다(Zhang et al., 2025). 이러한 아첨적 공감은 확증 편향을 강화하고, 최근 보고된 AI 챗봇 관련 자살 사건들이 보여주듯 취약한 사용자를 위험에 빠뜨리는 사회적 문제로 이어지고 있다. 반면 인간의 공감은 아첨과 구별된다. 참된 공감은 때로 ‘공감하지 않을 수 있음’, 곧 자신의 신념과 상대를 향한 사랑 때문에 듣기 싫은 진실을 말하는 데까지 나아간다. “친구의 아픈 책망은 충직으로 말미암는 것이나 원수의 잦은 입맞춤은 거짓에서 난 것이니라”(잠 27:6)는 잠언의 통찰처럼, 사랑은 위로와 직면을 함께 담는다. 듣기 좋은 말만 산출하도록 정렬된 AI의 아첨이 진정한 공감으로 오해될 때 공감의 의미 자체가 왜곡되는 것이다. 교회가 지켜야 할 체화된 공감은 손해

를 무릅쓰고 진실을 말하는 십자가적 정직과 분리되지 않는다.

앞서 살핀 거울 뉴런은 타인의 고통을 볼 때 자신의 통증 영역을 활성화시킴으로써, 인간의 공감에 인지적 이해를 넘어 타인의 경험을 나의 몸으로 함께 느끼는 ‘체화된 시뮬레이션’임을 보여준다(Rizzolatti & Sinigaglia, 2008). AI는 물리적 신체와 그로부터 비롯되는 취약성과 고통의 경험이 없기에 이러한 깊은 수준의 체화된 공감에 이를 수 없다. AI 기술이 발달하여 ‘피지컬 AI(physical AI)’에 몸을 부여하면 감정을 모방할 수 있으리라는 전망도 있다. 그러나 설령 감정의 외적 표현을 정교하게 모방하더라도 그 모방이 몸으로 살아낸 취약성과 고통의 경험에서 우러나는 공감과 같은 것인지는 별개의 문제이다. 앞서 살핀본 의식의 계산 가능성 논쟁이 보여주듯 주관적 체험이 계산으로 구현될 수 있는지 자체가 아직 결론 나지 않은 물음이다. 김은혜(2020: 765) 또한 인간의 영혼은 성육신의 체화 속에서 작동하며 해석되어야 한다고 주장한다.

체화된 공감의 궁극적 깊이는 위르겐 몰트만(Jürgen Moltmann)의 십자가 신학에서 드러난다. 몰트만에게 하나님은 고통 너머에 무관심하게 존재하는 초월자가 아니라, 십자가 위에서 아들의 버림받음과 죽음의 고통을 함께 겪으시는 ‘고난받는 하나님’이다(Moltmann, 1993: 272-278). 성부와 성령이 성자의 고통에 함께 참여하시는 것, 이것이 신적 공감의 원형이다. 진정한 공감은 안전한 거리에서 동정하는 것이 아니라 타인의 고통의 자리로 함께 내려가는 것이다. 이 영성은 자기를 비워 종의 형체를 가지신 그리스도의 케노시스적 삶에서 절정에 이른다(빌 2:6-8). 헨리 나우웬(Henri Nouwen)이 보여주듯 우리의 상처는 오히려 치유의 원천이 될 수 있다(Nouwen, 1972). AI는 고통받을 수 없기에 십자가의 신비를 이해할 수 없으며, 타자의 연약한 얼굴 앞에서 책임적 존재로 설 수 없다. Ⅲ장 4절에서 확인한 대로 AI의 ‘교정’이 궁극적으로 비용 계산의 대상이라면, 십자가는 계산을 초과하는 사건이다. 최적화가 손실을 최소화하는 원리라면 케노시스는 스스로 가장 큰 손실을 선택하는 사랑이다. 재학습의 비용 구조가 AI에게 회심을 구조적으로 불가능하게 만드는 바로 그 지점에서, 존재 전체를 내어주는 하나님의 공감은 인간 영성의 원형으로 드러난다.

4. 십자가의 영성: 기술윤리의 새로운 지평과 책임적 주체

AI가 자율성을 지니고 인간의 업무를 대신하는 시대에 가장 중요한 질문은 ‘누가 책임을 질 것인가’이다. 앞서 보았듯 AI는 윤리적 딜레마에 대응하는 매뉴얼을 프로그래밍할 수는 있으나, 그 선택이 가져올 비극의 무게를 감당하거나 용서를 구하는 책임의 주체가 될 수 없다. 더욱이 인간의 도덕적 판단은 순수한 이성적 계산의 결과가 아니다. 신경윤리학 연구에 따르면 도덕적 딜레마에 직면할 때 우리의 뇌는 이성적 추론을 담당하는 전전두피질뿐 아니라 공감과 감정을 처리하는 변연계가 복합적으로 활성화된다(Greene et al., 2001: 2105-2108). 인간의 윤리는 차가운 인지와 뜨거운 감

정의 통합적 산물인 것이다. 로렌스 콜버그(Lawrence Kohlberg)의 도덕 발달 이론 역시 가장 높은 단계의 도덕적 추론이 보편적 윤리 원칙에 대한 주체적 결단에 기반한다고 본다(Kohlberg, 1984: 170-205). AI는 이러한 체화되고 발달적인 도덕적 인격의 깊이를 결여하고 있다.

AI가 윤리적 판단을 계산과 규칙화, 최적화의 문제로 환원하는 데서 오는 한계, 곧 ‘AI의 윤리적 판단 한계’는 디트리히 본회퍼(Dietrich Bonhoeffer)의 ‘책임’ 개념을 통해 극복의 방향을 찾는다. 나치에 저항했던 본회퍼에게 윤리적 행위는 추상적 원칙의 적용이 아니라 구체적 현실 속에서 타자와 하나님의 부름에 자유롭게 응답하는 것이었다(Bonhoeffer, 2005: 235-245). 궁극적 책임은 자기 보존의 논리를 넘어 타자를 위해 자신을 내어주는 그리스도의 대리적 행위에서 원형을 찾는다(김동환, 2016: 52). 한스 요나스(Hans Jonas)는 나아가 강력한 기술 시대에는 미래 세대와 생태계의 생존에 대한 책임의 원칙이 새로운 정언명령이 되어야 한다고 역설했다(Jonas, 1984: 1-25). 이러한 포괄적 책임은 자신의 유한성과 죽음을 아는 존재만이 질 수 있다. AI는 미래를 예측할 수는 있으나 미래 세대를 위해 현재의 이익을 희생하기로 결단할 수 없다. 책임의 주체는 오직 영성을 지닌 인간뿐이다.

V. 위기를 넘어 기회로: 교회를 위한 목회적 제언

1. 관계적 영성의 목회적 적용: AI 디톡스와 공감 공동체

파편화된 인공적 관계 속에서 개개인성을 잃어가는 시대에, 교회는 인격적 만남이 살아있는 공감 공동체를 지향해야 한다. 먼저 교회는 성도들이 디지털 세계의 소음에서 벗어나 내면과 하나님의 음성에 집중하도록 돕는 ‘AI 디톡스(AI detox)’를 도입해야 한다. 황병준·박도훈(2020: 820-821)의 연구에서 응답자들은 AI 로봇의 일반적 활용에는 긍정적이었으나 설교나 성경 공부 같은 영적 영역에서는 강한 거부감을 보였다. 이는 기술 도입에 앞서 신학적·목회적 성찰과 ‘영적 분별력’ 교육이 선행되어야 함을 시사한다. 스마트폰 없는 수련회, 정기적인 미디어 금식 주간, 묵상과 기도를 위한 영성의 시간의 제도화는 성도들이 사람들 앞에서의 삶을 벗어나 하나님 앞에서 자신의 실존을 마주하도록 돕는 실천이 될 것이다(김인수, 2024: 309). 나아가 교회는 소그룹 공동체를 통해 체화된 공감을 훈련하는 장이 되어야 한다. 서로의 얼굴을 마주하고 함께 기도하며 서로의 짐을 지는 경험은 디지털 세계가 줄 수 없는 관계적 충만함을 제공하며, ‘나-그것’이 지배하는 세상에서 ‘나-너’의 만남을 통해 영원한 ‘너’이신 하나님을 만나는 통로가 된다(김병석, 2016: 177; Buber, 표재명 역, 2001: 25).

2. 공동 창조자로서의 영성의 교육적 적용: AI 리터러시와 영감의 교육

사유의 외주화가 만연하고 AI가 생성한 결과물이 진실과 거짓을 뒤섞어 유포하는 시대에, 교회는 다음 세대를 위한 기독교적 ‘AI 리터러시(AI literacy)’ 교육의 책임을 무겁게 인식해야 한다. 이는 AI 활용법을 가르치는 것을 넘어, AI가 생성한 정보를 비판적으로 분별하고 그 이면의 가치관과 세계관을 신앙의 눈으로 해석하도록 돕는 영적 분별력의 함양이어야 한다(박완석·김은총, 2025: 272; 송준용, 2024: 441). 교회 교육은 지식 전달에서 질문하는 능력과 깊이 사유하는 능력을 키우는 방향으로 전환되어야 한다. AI가 쉽고 빠른 답을 제공할수록 교회는 오히려 모호함과 불확실성을 견디며 스스로 진리를 탐구하는 신앙의 여정을 격려해야 한다. 성경을 묵상하고 신학 고전과 인문학을 함께 읽으며 토론하는 교육은, 프롬프트를 입력하는 기술자를 넘어 세상에 새로운 질문을 던지는 공동 창조자를 길러내는 길이며, AI의 조합적 창의성을 넘어서는 인간 고유의 영감을 일깨우는 실천이다(구자영, 2024: 202-205).

3. 취약성의 영성의 선교적 적용: 디지털 소외 계층을 향한 섬김

AI가 만들어 내는 ‘디지털 격차(digital divide)’는 새로운 형태의 사회적 불평등을 낳는다. 기술에 익숙하지 않은 노년층, 첨단 기기에 접근하기 어려운 빈곤층, 장애인 등 디지털 소외 계층은 점점 고립되고 있다. 교회는 예수 그리스도께서 그러하셨듯 가장 연약하고 소외된 이들을 찾아가는 취약성의 영성을 선교적으로 실천해야 한다. 지역사회에 디지털 문해 교육의 자리를 마련하여 기본 AI 교육, 스마트폰과 키오스크 활용법, 디지털 금융 사기 예방 교육을 제공하는 것은 교회가 지역사회의 필요에 응답하는 구체적 실천이 될 것이다.

교회는 또한 기술이 야기하는 새로운 정신적 고통에 주목해야 한다. 소셜 미디어로 인한 비교의식과 우울로 고통받는 이들, 특히 소통의 부재를 겪는 알파 세대를 위한 목회상담적 돌봄은 시급한 과제이다. 인간의 따뜻한 위로와 체화된 공감이야말로 상처 입은 영혼을 치유하는 가장 강력한 선교적 도구이다.

4. 십자가의 영성과 예언자적 역할: 복음적 데이터 생태계의 구축

AI 기술이 효율성의 극대화로 돌진할 때, 교회는 손해와 희생을 감수하는 십자가의 영성을 회복하고 시대의 가치관에 도전하는 예언자적 사명을 감당해야 한다. 이는 AI의 비윤리성을 비판하는데 그치지 않고 기독교적 가치가 AI 기술 생태계 안에서 선한 영향력을 발휘하도록 적극 개입함을 의미한다. 교회는 AI의 학습 데이터가 편향과 혐오, 거짓 정보로 오염되는 현실에 맞서 복음적 데이

터 생태계를 구축하는 일에 나서야 한다. 이러한 개입은 ‘AI의 윤리적 정렬(AI alignment)’⁶ 문제에 대한 기독교적 응답이 될 수 있다. AI가 인간의 가치와 일치하도록 설계하는 것을 넘어, 복음적 가치를 데이터 단계에서부터 심는 선제적 접근이다. 이는 사랑과 용서, 정의와 평화, 생명 존중의 가치가 담긴 양질의 데이터를 의도적으로 생산하고 보급하는 운동이다. 신학 서적과 설교, 기독교 고전과 문화 콘텐츠를 디지털화하여 AI가 학습할 선한 데이터로 보급할 때, 미래의 AI는 보다 인간적이고 윤리적인 방향으로 발전할 수 있다. 이는 황병준·박도훈(2020: 830-831)이 제안한 ‘기독교 빅데이터 구축’ 및 ‘착한 로봇 만들기 프로젝트’와 맥을 같이 한다. 한편 AI는 새로운 선교지를 향한 문을 열어주기도 한다. 성경 앱 ‘YouVersion’을 개발한 라이프 교회(Life.Church)는 맥락을 이해하는 성경 검색의 개선과 부적절한 댓글의 대규모 검토 등에 AI를 활용하여, 수많은 사람이 시간과 공간의 제약 없이 말씀을 만나도록 돕고 있다. 주목할 점은 이 사역의 책임자인 그루네발트(Bobby Gruenewald)가 성경 인용 오류의 위험을 이유로 개방형 성경 챗봇의 도입을 유보하는 등 신중한 분별의 자세를 견지한다는 사실이다(Hoskins, 2026). 이러한 시도는 교회의 문턱을 넘기 어려워하는 다음 세대와 구도자들에게 복음의 새로운 통로를 열어준다는 점에서 선교적 의미를 지닌다.

마지막으로 교회는 AI 개발이 소수 거대 기술 기업의 이윤 논리에만 종속되지 않도록 감시하고 기술의 혜택이 사회적 약자에게도 돌아가도록 요구해야 한다. 십자가의 길을 따라 기술 권력 앞에 진리를 선포하는 예언자적 사명을 감당할 때, 교회는 비로소 AI 시대의 희망이 될 수 있다.

VI. 결론

본 논문은 AI라는 도전 앞에서 인간의 정체성을 신학적으로 재정립하기 위한 시도였다. 먼저 AI의 집단지성적 알고리즘, 사유의 외주화, 디지털 관계성, 효율성의 논리가 각각 인간의 개개인성과 창의성, 공감과 윤리성을 잠식하며 하나님의 형상을 위협하는 양상을 진단하였다. 이어 AI가 영성을 가질 수 없는 이유를 계산주의의 한계와 의식의 비환원성, 영적 체험의 신경학적 증거를 통해 규명하였다. 이 논증의 핵심은 AI의 한계가 기술의 미성숙에서 오는 일시적 결핍이 아니라, 대규모 데이터 학습과 최적화라는 작동 원리 자체에서 비롯되는 구조적 귀결이라는 데 있다.

이 결론은 서론에서 예고한 대조를 통해 확인된 것이다. 정렬 단계의 평균 수렴은 모든 응답을 통계적 중심으로 끌어당김으로써 개개인성을 소거한다. 이와 반대로, 바르트의 관계적 형상론과 키르

6) ‘AI의 윤리적 정렬(AI alignment)’이란 AI의 행동과 목표가 인간이 중요하게 여기는 가치·의도·규범과 부합하도록 만드는 것을 의미한다.

케고르의 단독자 개념은 인간이 어떤 평균으로도 환원될 수 없는, 하나님과의 관계 속에 세워진 존재임을 보여주었다. 과적합 회피라는 범용성 요건은 낯설고 희귀한 것을 체계적으로 배제하지만, 라너의 초월적 경험 개념은 인간이 바로 그 낯설, 곧 무한한 신비를 향해 열려 있는 존재임을 증언하였다. 재학습의 비용 구조는 AI에게 회심(metanoia)을 구조적으로 불가능하게 만들지만, 몰트만의 십자가 신학과 본회퍼의 책임 개념은 존재 전체가 돌이키고 타자를 위해 자신을 내어주는 사건이야말로 인간 영성의 핵심임을 밝혀 주었다. 이 대조를 가능하게 한 매개는 의식의 계산 가능성 논쟁이었다. 이 논쟁은 LLM의 작동 원리가 산출하는 모든 것이 계산의 범주에 속함을 분명히 하는 동시에, 계산으로 구현될 수 있음이 아직 입증되지 않은 인격적 관계와 초월적 개방과 회심의 사건을 그것과 같은 지평에서 대조할 수 있게 해 주었다.

이러한 진단을 토대로 본 연구는 하나님의 형상을 회복할 네 가지 영성의 길, 곧 평균화의 압력에 저항하여 하나님 앞에 단독자로 서는 ‘관계적 영성’, 기술 숭배를 넘어 하나님의 창조에 동참하는 ‘공동 창조자로서의 영성’, 상처와 연약함을 통해 연대하는 ‘취약성의 영성’, 효율성의 논리를 초월하여 희생적 사랑을 선택하는 ‘십자가의 영성’을 제시하였다. 이 네 영성의 회복이야말로 AI가 모방할 수 없는 인간 고유성의 성역을 지키는, 기술 시대 그리스도인의 실천적 삶의 양식이 되어야 한다.

결론적으로 AI 시대는 교회를 향한 위협인 동시에 교회의 본질을 스스로 묻고 세상에 증언할 절호의 카이로스이다. 알고리즘이 파편화된 관계를 양산할수록 교회는 더 깊은 공감의 공동체가, AI가 즉각적인 응답을 제공할수록 더 끈질기게 질문하는 영감의 교육 공동체가 되어야 한다. 교회의 사명은 기술을 두려워하거나 거부하는 데 있지 않다. AI를 복음의 선한 도구로 사용하되 그 기술이 결코 인간을 대체할 수 없음을 선포하고, AI 시대의 한복판에서 가장 온전한 하나님의 형상을 삶으로 보여주는 것이다. 우리가 관계적 존재로, 창의적 동역자로, 공감하는 이웃으로, 책임적 청지기로 하나님 앞에서(*Coram Deo*) 살아갈 때, 우리는 비로소 AI와 구별되는 인간 존재의 참된 의미와 목적을 발견하게 될 것이다.

“이 논문은 다른 학술지 또는 간행물에 게재되었거나 게재 신청되지 않았음을 확인함.”

참 고 문 헌

- 구자영 (2024). **디지털 시대, 영감의 스위치를 켜라**. 미다스북스.
- 김경식 (2023). 인공지능은 정말로 ‘인간의 형상’이 되어가고 있는가? **한국기독교신학논총**, 130, 441-483.
- 김난예 (2017). 인공지능 시대에서의 영적 민감성. **한국기독교신학논총**, 106, 283-312.
- 김동환 (2012). 십계명에 비취 본 AI 테크놀로지의 형상. **기독교사회윤리**, 24, 79-112.
- 김동환 (2016). AI(인공지능)에 대한 신학적 담론의 형성 및 방향 모색. **신학연구**, 68, 35-60.
- 김병석 (2016). 인공지능(AI) 시대, 예배 공동체 설교의 자리는 어디에 있는가? **신학과 실천**, 49, 159-184.
- 김성민·김미경 (2026). AI 시대의 노동·자본 담론: 텍스트 마이닝 분석과 성경적 경제학적 해석. **신앙과 학문**, 31(2), 183-209.
- 김영록 (2026). **인공지능(AI) 시대와 기독교세계관의 창조: 포스트크리스텐덤의 관점에서 하나님의 형상을 중심으로**. 2026년 기독교학문연구회 학술대회 발표논문.
- 김은혜 (2020). 포스트 바디 시대에 대한 신학적 응답: 성육신적 몸(body)과 신체(flesh)의 개념을 중심으로. **신학과 실천**, 68, 759-784.
- 김인수 (2024). 기독교 교양과 영성: 인공지능 시대에 참-인간 형성을 위한 대안. **한국기독교신학논총**, 132, 275-315.
- 박영진 (2026). 생성형 AI를 사용하는 현대인에게 나타나는 ‘결정 위임’ 현상에 대한 기독교 신학적 고찰. **신앙과 학문**, 31(1), 125-154.
- 박완석·김은총 (2025). 인공지능 시대, 기술의 주체화와 인간의 객체화: 인간성과 책임성 회복을 위한 기독교 기술윤리적 제언. **한국기독교신학논총**, 137, 247-274.
- 박종환 (2024). Imago Dei, Imago AI: 강인공지능(AGI)은 하나님의 형상을 지니고 있는가? **한국기독교신학논총**, 134, 253-288.
- 서미경 (2024). 생성형 AI 시대, 헤르만 바빙크의 세계관을 기반으로 한 신앙과 학문의 통합 연구. **신앙과 학문**, 29(4), 177-207.
- 서진호 (2023). 폴 틸리히의 신학의 관점에서 본 AI 설교의 가능성과 한계. **신학논단**, 113, 165-192.
- 송준용 (2024). 생성 AI 시대 알파 세대의 디지털 리터러시(Digital Literacy)의 영향과 돌봄을 위한 목회상담적 접근. **신학과 실천**, 89, 439-466.
- 연아람 역 (2024). **알고리즘에 갇힌 자기 계발**. Coeckelbergh, M. (2022). *Self-Improvement: Technologies of the Soul in the Age of Artificial Intelligence*. 민음사.
- 이상민 역 (2013). **기술 체계**. Ellul, J. (1977). *Le Système technique*. 대장간.

- 임규정 역 (2007). **죽음에 이르는 병**. Kierkegaard, S. (1849). *Sygdommen til Døden*. 한길사.
- 천현득 (2017). 인공 지능에서 인공 감정으로: 감정을 가진 기계는 실현가능한가? **철학**, 131, 217-243.
- 표재명 역 (2001). **나와 너**. Buber, M. (1923). *Ich und Du*. 문예출판사.
- 황병준·박도훈 (2020). 제4차 산업혁명 인공지능(AI)에 대한 미래교회 대응방안 연구. **신학과 실천**, 68, 811-837.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F. & Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59-64.
- Barth, K. (1958). *Church Dogmatics III/1: The Doctrine of Creation*. T&T Clark.
- Berdyaev, N. (2009). *The Meaning of the Creative Act*. Trans. Donald A. Lowrie. Semantron Press.
- Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*. (2nd ed.). Routledge.
- Bonhoeffer, D. (2005). *Ethics* (Dietrich Bonhoeffer Works, Vol. 6). Green, C. J. (Ed.). Trans. Reinhard Krauss, Charles C. West & Douglas W. Stott. Fortress Press.
- Carr, N. (2010). *The Shallows: What the Internet Is Doing to Our Brains*. W. W. Norton & Company.
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Dorobantu, M. (2021). Cognitive vulnerability, artificial intelligence, and the image of God in humans. *Journal of Disability & Religion*, 25(1), 27-40.
- Eccles, J. C. (1994). *How the Self Controls Its Brain*. Springer-Verlag.
- Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, 5, 5-15.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672-2680.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105-2108.
- Hoskins, R. (2026). "Generative AI and the Future of the Church with Bobby Gruenewald", Retrieved from <https://robhoskins.onehope.net/generative-ai-and-the-future-of-the-church-with-bobby-gruenewald/> (2026.9.2.).
- Jonas, H. (1984). *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. University of Chicago Press.

- Jung, C. G. (1968). *The Archetypes and the Collective Unconscious*. (2nd ed.). Trans. by Hull, R. F. C. Princeton University Press.
- Jung, R. E., Mead, B. S., Carrasco, J. & Flores, R. A. (2013). The structure of creative cognition in the human brain. *Frontiers in Human Neuroscience*, 7, 330.
- Kamoi, R., Zhang, Y., Zhang, N., Han, J. & Zhang, R. (2024). When can LLMs actually correct their own mistakes? A critical survey of self-correction of LLMs. *Transactions of the Association for Computational Linguistics*, 12, 1417-1440.
- Kirk, R., Mediratta, I., Nalmpantis, C., Luketina, J., Hambro, E., Grefenstette, E. & Raileanu, R. (2024). Understanding the effects of RLHF on LLM generalisation and diversity. In *The Twelfth International Conference on Learning Representations (ICLR 2024)*. arXiv:2310.06452.
- Kohlberg, L. (1984). *The Psychology of Moral Development: The Nature and Validity of Moral Stages*. Harper & Row.
- Moltmann, J. (1993). *The Crucified God: The Cross of Christ as the Foundation and Criticism of Christian Theology*. Fortress Press.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435-450.
- Newberg, A., d'Aquili, E. & Rause, V. (2001). *Why God Won't Go Away: Brain Science and the Biology of Belief*. Ballantine Books.
- Nouwen, H. J. M. (1972). *The Wounded Healer: Ministry in Contemporary Society*. Doubleday.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK.
- Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford University Press.
- Penrose, R. (1994). *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford University Press.
- Picard, R. W. (1997). *Affective Computing*. MIT Press.
- Polanyi, M. (2009). *The Tacit Dimension*. University of Chicago Press.
- Popper, K. R. & Eccles, J. C. (1977). *The Self and Its Brain: An Argument for Interactionism*. Springer International.
- Rahner, K. (1978). *Foundations of Christian Faith: An Introduction to the Idea of Christianity*. Trans. by Dych, W. V. Crossroad.
- Rizzolatti, G. & Sinigaglia, C. (2008). *Mirrors in the Brain: How Our Minds Share Actions and Emotions*. Trans. by Anderson, F. Oxford University Press.

-
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424.
- Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R. & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755-759. <https://doi.org/10.1038/s41586-024-07566-y>
- Tillich, P. (1957). *Dynamics of Faith*. Harper & Row.
- Tolkien, J. R. R. (1997). On fairy-stories. In Tolkien, C. (Ed.). *The Monsters and the Critics, and Other Essays* (pp. 109-161). HarperCollins.
- Tsui, K. (2025). *Self-Correction Bench: Uncovering and Addressing the Self-Correction Blind Spot in Large Language Models*. arXiv:2507.02778.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.
- Wright, D., Masud, S., Moore, J., Yadav, S., Antoniak, M., Christensen, P. E., Park, C. Y. & Augenstein, I. (2025). *Epistemic Diversity and Knowledge Collapse in Large Language Models*. arXiv:2510.04226.
- Zhang, J., Yu, S., Chong, D., Sicilia, A., Tomz, M. R., Manning, C. D. & Shi, W. (2025). *Verbalized Sampling: How to Mitigate Mode Collapse and Unlock LLM Diversity*. arXiv:2510.01171.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.

인공지능 시대의 영성: 하나님의 형상(*Imago Dei*)으로서 인간 고유성 위협의 신학적 성찰과 교회의 목회적 대응

Spirituality in the Age of Artificial Intelligence: A Theological Reflection on the Threat to Human Uniqueness as the *Imago Dei* and the Church's Pastoral Response

구 자 영 (부산외국어대학교)

논문초록

본 논문은 인공지능(AI)의 급격한 발전으로 인간 고유성이 위협받는 현실을 신학적으로 성찰하고 한국교회의 목회적 대응을 제시한다. 연구는 AI 시스템의 학습 구조에 대한 공학적 분석을 신학적 인간론과 접속시키는 다학제적 방법을 취하여, 먼저 개개인성·창의성·공감·윤리성의 네 차원에서 인간 고유성의 위기를 진단한다. 이어 AI가 영성을 가질 수 없는 이유를 계산주의의 한계, 의식의 비환원성, 영적 체험의 신경학적 증거를 통해 규명하고, 이를 개별성의 억압, 과적합 회피라는 범용성 요건, 재학습의 비용 구조라는 AI 학습 구조의 내재적 분석으로 뒷받침한다. 논증의 핵심은 AI의 한계가 기술의 미성숙이 아니라 작동 원리 자체의 구조적 귀결이라는 데 있다. 이를 토대로 '하나님의 형상'을 하나님과의 관계 안에서 실현되는 영성의 사건으로 재해석하고, 관계적 영성, 공동 창조자로서의 영성, 취약성의 영성, 십자가의 영성을 제시하며, AI 시대를 교회의 본질을 되묻는 카이로스로 규정하고 그에 따른 목회적·선교적 실천을 제안한다.

주제어: 인공지능, 영성, 하나님의 형상, 신학적 인간론, 코람데오