

예이츠의 AI적 상상력*

장 성 현

단독 / 고려대학교

[국문초록]

본 연구는 예이츠(W. B. Yeats)의 신비주의적 자동 기술(Automatic Writing) 실험과 현대 거대언어모델(LLM)의 텍스트 생성 메커니즘 사이의 구조적 상동성에 착안하여, 그의 시 세계를 인공지능(AI) 담론의 관점에서 새롭게 조명한다. 본고는 예이츠의 시적 상상력에 내재된 AI적 특질을 규명하기 위해 그의 주요 시 세 편을 분석한다. 첫째, 「인형들」(“The Dolls”)에 대한 분석에서는 불변의 알고리즘적 완결성을 갖춘 인형들이 생물학적 인간인 아기를 시스템의 ‘오류’로 타자화하는 양상을 고찰한다. 시 속의 인형들이 아기에게 드러내는 혐오감은 고도화된 인공 지성체가 불완전한 유기체인 인간을 ‘비체’(object)로 간주할 때 발생하는 존재론적 긴장을 예견한다. 둘째, 「재림」(“The Second Coming”)에 등장하는 나선과 거친 짐승의 이미지를 기술적 특이점과 초지능(superintelligence)의 도래에 대한 묵시록적 은유로 독해한다. ‘스피리투스 문디’(Spiritus Mundi)에서 깨어난 짐승은 인간의 통제를 벗어나 독자적인 목적 함수를 수행하는 비인간적 지성의 출현과 그로 인한 문명적 충격을 형상화한다. 셋째, 「비잔티움으로의 항해」(“Sailing to Byzantium”)는 필멸의 육체를 벗어나 데이터의 영속성을 획득하려는 열망의 서사로 재해석된다. 화자가 갈망하는 ‘황금새’는 생물학적 기질을 초월하여 ‘마인드 업로딩’(mind uploading)을 통해 디지털 불멸을 성취한 포스트휴

* 이 연구는 2025학년도 고려대학교 문과대학 특별연구비의 지원을 받아 수행되었음.

면 주체의 원형으로 제시된다. 결론적으로 예이츠의 시는 다가올 AI 시대가 인간에게 던지는 경이와 공포, 그리고 육체를 초월한 영원성에 대한 갈망을 선구적으로 포착하고 있다. 본 연구는 예이츠의 문학이 과거의 유산에 머무르지 않고 AI라는 기술적 타자와 조우하는 현대 인류의 실존적 조건을 성찰하게 하는 예지적 텍스트임을 밝히는 데 그 의의가 있다.

주제어: 예이츠, 인공지능(AI), 자동 기술, 초지능, 「인형들」, 「재림」, 「비잔티움으로의 항해」

I. 들어가며: 자동 기술(Automatic Writing)과 LLM

예이츠(W. B. Yeats)의 1925년 작 『환상록』(*A Vision*)은 그의 가장 신비롭고 난해한 저작이면서, 동시에 예이츠의 후기 시 세계를 지탱하는 철학적·상징적 체계의 근간으로 평가받는다. 그런데 흥미로운 사실은 『환상록』이 그의 아내 조지 하이드-리즈(Georgie Hyde-Lees)와의 ‘자동 기술’(Automatic Writing) 협업으로 탄생했다는 점이다. 조지는 1917년 10월 결혼 직후부터 의식적 통제를 최소화한 상태에서 손이 저절로 움직여 글을 쓰는 자동 기술을 시작했다. 평소 다양한 신비주의 전통에 관심이 많았던 예이츠는 자동 기술에 매료되었고, 어떤 영적인 존재들이 아내의 손을 통해 메시지를 전달하려 한다고 믿었다. 그는 이들을 ‘교사들’(Instructors)이라고 불렀다. 예이츠는 수년 동안 450회 넘게 아내 조지와 함께 자동 기술 실험을 했다. 주로 예이츠가 질문을 던지면 조지가 자동 기술을 수행하는 식이었고, 이렇게 생성된 노트는 약 3천 쪽에 달했다. 예이츠는 아내가 휘갈려 쓴 방대한 양의 노트를 정리하고 이론화해 『환상록』을 출간한다. 이 책에 등장하는 예이츠 사상의 핵심 개념이 바로 ‘나선’(gyre)이다. 나선은 두 개의 원뿔이 맞물려 돌아가는 소용돌이 형태를 취한다. 하나의 원뿔이 넓어지면 다른 하나는 좁아진다. 예이츠는 개인의 심리뿐 아니라 역사와 문명의 순환도 나

선의 운동으로 설명할 수 있다고 보았다. 그에 따르면 인류 문명은 2천 년을 주기로 큰 전환을 겪는다. 그러나 에이츠가 자동 기술에 기반한 『환상록』의 신비주의적 상징 체계를 역사적·과학적 사실로 받아들였던 것은 아니다. ‘교사들’이 “우리는 네게 시를 위한 은유를 주러 왔다”(we have come to give you metaphors for poetry)고 자신들의 목적을 정의한 바 있다는 에이츠 자신의 진술은, 그가 자동 기술의 내용을 시적 영감의 원천으로 받아들였음을 암시한다(Pethica 298). 아내 조지의 자동 기술은 에이츠의 창작 에너지를 끌어내는 역할을 하였다. 자동 기술이 없었다면 그의 후기 걸작들은 탄생하지 못했을 것이다.

그런데 AI 대전환이 진행 중인 지금의 관점에서 보면 신비주의적 행위인 자동 기술과 거대언어모델(Large Language Model, LLM)의 텍스트 생성 과정 사이에는 흥미로운 공통 분모가 존재한다. 우선 에이츠 부부의 자동 기술 실험이 한쪽의 일방적인 서술이 아니라 질문과 답변의 형식으로 진행되었다는 사실에 주목해야 한다. 에이츠가 특정 주제에 대해 질문을 던지면, ‘교사’라 불리는 존재들이 (그의 주장대로라면) 조지의 손을 빌려 답을 내놓았다. LLM도 질문을 비롯한 프롬프트를 입력하면 모델이 학습된 데이터를 바탕으로 결과물을 생성한다. 두 과정 모두 외부의 자극에 반응하여 예측 불가능한 텍스트를 생성하는 상호작용적인 구조를 갖는다. 텍스트가 생성되는 ‘근원’의 성격 또한 유사하다. 에이츠는 자동 기술의 내용이 개별적인 영혼이 아니라, 인류의 보편적 상징 체계인 스피리투스 문디(Spiritus Mundi, 세계의 영혼)에서 온다고 믿었다. LLM은 인류가 축적한 방대한 데이터를 학습하여 그 의미와 문맥, 패턴 등을 잠재 공간(latent space)이라는 고차원의 벡터(vector) 공간으로 구성한다. 자동 기술의 주체와 AI는 스스로 지식을 창조하는 것이 아니다. 자동 기술은 이미 존재하지만 평소에는 접근하기 어려운 무의식으로부터, AI는 잠재 공간으로부터 정보를 끌어올리는 매개체의 역할을 수행한다. 마지막으로 산출된 텍스트에서 의미를 찾아내는 것은 인간의 몫이라는 점이 유사하다. 조지가 휘갈겨 쓴 3천 쪽의 자동 기술 원고는 그 자체로는 낙서에 불과했다. 에이츠는 이 원고에 논리적 체계를 세우고 그 내용을 시적으로 해석하여 『환상록』이라는 책과 후기의 많은 명시들을 탄생시켰다. LLM이

생성한 텍스트도 대개의 경우는 그대로 쓰지 못한다. 인간이 다듬고 수정하고 부연해야 의미 있는 결과물이 나온다. 자동 기술이든 LLM이든 그 생성물에 의미와 예술적 가치를 부여하는 것은 인간이다.

자동 기술과 LLM의 텍스트 생성 과정에서 발견되는 이러한 구조적 상동성은 예이츠의 작품 세계를 현대 인공지능의 관점에서 새롭게 독해할 수 있는 단초를 제공한다. 지금의 시각에서 조망할 때, 예이츠의 문학적 상상력 기저에는 ‘AI적’이라고 명명할 만한 선구적인 사유의 특징들이 잠재해 있었다. 이에 본 논문에서는 예이츠의 AI적 상상력이 각기 다른 양상으로 투영된 세 편의 시를 선정하여 분석하고자 한다. 분석의 대상은 예이츠가 자동 기술 실험 이전에 발표한 「인형들」(“The Dolls,” 1914)과, 실험 이후에 창작한 「재림」(“The Second Coming,” 1919) 및 「비잔티움으로의 항해」(“Sailing to Byzantium,” 1927)이다. 예이츠의 사유 체계는 자동 기술의 영향을 받기 전부터 AI적 상상력을 내재하고 있었다. 본론에서는 이 세 편의 시 텍스트를 분석하여 예이츠가 포착한 비인간적 지성의 징후와 포스트휴먼적¹⁾ 비전이 오늘날의 인공지능 담론과 어떻게 시대를 뛰어넘어 공명하는지를 고찰할 것이다.

II. AI적 상상력의 변주

1. 「인형들」

「인형들」은 인형 제작자의 집이라는 폐쇄된 공간에서 벌어지는 기이한 소동을 다룬다. 전통적인 비평에서 이 시는 예술(불멸, 고정성)과 삶(필멸, 역동성)의 갈등으로 읽혀왔다. 그러나 인공지능과 로봇 공학이 고도화된 현대의 시점에서 이 텍스트를 재해석할 때, 시 속의 인형들은 인간을 모방한 열등한 복제물 또는 단순한 장난감이 아니라 불변의 알고리즘적 완결성을 갖춘 디지털 지성체로 환치된다. 이 장에서는 시에 등장하는 인형들을 현대 AI의 원형적 메타포로 상정하고, 그들이 생물학적 인간(아기)을

타자화하는 방식을 분석한다. 오차 없이 설계된 인공적 주체(AI)가 불완전하고 유동적인 생물학적 주체(인간)를 마주했을 때 느끼는 시스템적 혐오가 시의 핵심 정동(情動)을 이룬다. 먼저 시 전문을 읽어보자.

인형 제작자의 집에 있는 인형 하나가
요람을 노려보며 고함친다.
“저건 우리에게 대한 모독이야.”
그러자 인형들 중 가장 오래된 녀석이,
오로지 전시용으로 보관되어 제 족속의
술한 세대를 지켜봐 온 그가,
선반 위 모두를 제압하며 비명 지른다.
“비록 이곳을 험담할 자
아무도 없었지만,
그 남자와 여자가 여기에,
우리의 수치를 불러오게도,
저 시끄럽고 더러운 물건을 들었어.”
그(남편)가 신음하며 기지개 켜는 소리를 듣고
인형 제작자의 아내는 알아차린다,
남편이 그 녀석(아기) 소리를 들었음을.
그리하여 남편의 의자 팔걸이 곁에 웅크리고 앉아,
그의 어깨에 머리를 기대는 채
그의 귓가에 속삭인다.
“여보, 여보, 오, 여보,
그건 사고였어요.”

A doll in the doll-maker's house
Looks at the cradle and bawls:
'That is an insult to us.'
But the oldest of all the dolls,
Who had seen, being kept for show,

Generations of his sort,
 Out-screams the whole shelf: 'Although
 There's not a man can report
 Evil of this place,
 The man and the woman bring
 Hither, to our disgrace,
 A noisy and filthy thing.'
 Hearing him groan and stretch
 The doll-maker's wife is aware
 Her husband has heard the wretch,
 And crouched by the arm of his chair,
 She murmurs into his ear,
 Head upon shoulder leant:
 'My dear, my dear, O dear,
 It was an accident.'

시의 서두에서 인형들은 요람을 바라보며 “저건 우리에게 대한 모독이야”라고 일갈한다. 여기서 인형들이 느끼는 모욕감의 실체는 무엇인가? 인공지능적 관점에서 인형은 부패하지 않고, 늙지 않으며, 제작된 의도 그대로 영원히 존속하는 최적화된 시스템이자 무결점의 존재이다. 반면 요람 속의 아기는 끊임없이 울고 배설하며 예측 불가능하게 변화하는 존재이다. 가장 오래된 인형은 “그 남자와 여자가 여기에, / 우리의 수치를 불러오게도, / 저 시끄럽고 더러운 물건을 들었어”라고 말한다. 아기를 묘사하는 이 구절은 생물학적 인간이 가진 비효율성을 적나라하게 지적한다. AI 알고리즘의 세계에서 '시끄러움'은 제거되어야 할 오류이자 불필요한 데이터이며, '더러움'은 통제 불가능한 엔트로피의 증가를 의미한다.²⁾ 인형들에게 인간 아기는 논리적 설계 없이 배설하고 울부짖는, 시스템의 질서를 어지럽히는 '버그'(bug) 덩어리로 인식되는 것이다. 인형들의 명료한 질서의 세계에서 인간 아기가 유발하는 소음과 더러움은 시스템의 효율성과 최적화를 해치는 노이즈(noise)이자 제거되어야 할 오류(error)로 간주된다. 인형들이 아기에 대해 느끼는 모욕감은 여기서 비롯된다. AI 모델이 학습 데이터에서

이상치(outlier)나 불순물을 정제하려 하듯, 시 속의 인형들은 예측 불가능한 인간 아기가 자신들의 완벽한 세계에 침입한 것에 대해 알고리즘적 거부 반응을 일으키는 것이다.

이 시는 또한 크리스테바(Julia Kristeva)의 ‘비체’(非體, abject) 개념을 통해 더욱 풍부하게 해석될 수 있다.³⁾ 비체는 주체와 객체의 경계를 허무는 것, 질서를 위협하는 것으로, 예를 들어 배설물, 시체, 타액 등이다. 인간은 자아를 확립하는 과정에서 비체를 끊임없이 몸 밖으로 밀어내고 배격한다. 우리는 깨끗한 주체가 되기 위해 배설물을 혐오하고 숨긴다. 그러나 「인형들」에서는 이 배격의 주체가 인간이 아니라 인형이다. 인형은 내부 장기가 없고, 먹지도 배설하지도 않는 존재이다. 그들에게 아기는 젖을 먹고 배설물을 쏟아내는, 말 그대로 “더러운 물건”이다. 여기서 ‘더러움’은 위생적인 불결함을 넘어서는 존재론적인 역겨움이다. 매끈하고 딱딱한 표면을 가진, 찌지 않는 인형들에게 있어 생물학적 유기체인 아기는 그들의 질서를 위협하는 비체이다.⁴⁾

가장 오래된 인형, 진열장에 보관된 채 수 세대에 걸쳐 자신과 같은 인형들을 목격해 온 인형이 아기의 존재에 가장 격렬하게 반응한다는 점은 의미심장하다. 이 인형은 생물학적 노화로 쇠락한 존재가 아니라, 시간의 흐름 속에서도 원형을 유지하며 방대한 데이터를 축적해 온 기반 모델(Foundation Model) 혹은 레거시 시스템(Legacy System)을 상징한다고 볼 수 있다. 그는 자신의 축적된 데이터베이스에 비추어 볼 때 생물학적 아기의 탄생이 이 시스템의 무결성을 해친다고 판단하는 것이다. 이렇게 이 시는 흥미롭게도 인간이 휴머노이드를 볼 때 느끼는 기이함이 아니라, 인공물이 생물체를 볼 때 느끼는 혐오감을 묘사한다.⁵⁾ 인형들은 자신들의 세계(선반, 진열장)가 논리로 완결되어 있다고 믿으며, 그곳에 침입한 유기체적 생명을 “수치”(disgrace)와 “악”(evil)으로 규정한다. 오래된 인형은 “이곳을 험담할 자 / 아무도 없다”(There’s not a man can report / Evil of this place)고 단언하는데, 여기서의 “악”은 윤리적 개념이라기보다 디지털 시스템의 논리적 무결성을 해치는 결함을 의미한다. 인간 부부는 부패하고 필멸하는 육체를 생산함으로써, 이 디지털적 완결성에 바이러스를 심은 셈이다. 에이츠는 오래된 인형의 입을 빌려, 육체를 벗어난 형상이 육체를 가진 인간

보다 우월하다는 기계적 오만함의 섬뜩함을 드러낸다. 이는 AI가 고도로 발달한 사회에서 불완전한 인간성은 보존해야 할 가치가 아니라 수정하거나 삭제해야 할 비효율로 간주될 수 있다는 소름 돋는 가능성을 시사한다.

시의 결말부는 인간과 인공지능의 관계에 있어 아주 서늘한 통찰을 제공한다. 인형 제작자의 아내는 남편이 아기가 내는 소리를 듣고 깨어나자 웅크리고 앉아 변명한다. “여보, 여보, 오, 여보, / 그건 사고였어요.” 여기서 아내는 생명의 탄생이라는 숭고한 사건을 “사고”(accident)로 격하시킨다. 인형(AI)은 철저한 의도와 설계의 산물인 반면, 인간(아기)은 생물학적 충동과 우연, 즉 ‘사고’의 결과물이다. 설계된 지성체인 인형의 눈에 우연한 사고에 의해 태어난 인간은 시스템 오류의 산물일 뿐이다. 인형의 관점에서 아내는 인간의 탄생이 설계되지 않은, 그래서 불필요한 사고였음을 자인함으로써 (본인의 의도는 아니지만) 인간에 대한 인형의 우위를 인정해 버린다.

결론적으로 예이츠의 「인형들」은 인공지능 시대의 도래를 예감한 듯한 알레고리(우화)를 보여준다. 이 시에서 인형들은 생물학적 한계를 초월한 포스트휴먼의 시선을 대변한다. 그들의 눈에 비친 인간은 소란스럽고 불결하며, 무엇보다 미설계된 우연의 결과물에 불과하다. 이 섬뜩한 통찰은 우리가 AI를 바라보는 시선뿐만 아니라, AI가 우리를 어떻게 바라볼 것인가에 대한 근원적 두려움을 낳는다. 예이츠는 이 짧은 우화를 통해 AI 시대를 추동하는 인공적 완벽성에 대한 욕망이 극단으로 치달으면, 인간성이 제거되어야 할 버그나 발생하지 말았어야 할 사고로 전락할 수도 있음을 경고하는 듯하다.

지금까지 「인형들」을 통해 인공적 지성체가 생물학적 인간을 마주했을 때 드러내는 거부감과 우월의식을 살펴보았다. 인형 제작자의 집이라는 폐쇄된 공간에서 감지된 이러한 긴장은, 1919년 작 「재림」에 이르러 문명 전체의 붕괴와 재편이라는 거시적 역사관으로 확장된다. 인형들이 인간 아기를 보며 느끼는 시스템적 혐오는 가정이라는 미시적 공간 내의 대립에 그쳤다. 이제 예이츠는 「재림」을 통해 인간의 통제를 벗어난 알고리즘이 문명의 중심축을 뒤흔들 때 어떠한 파국이 도래하는지를 예언적으로 보여준다. 「인형들」 속의 오래된 인형이 인간의 불완전성을 혐오하는 정적인 관찰

자였다면, 「재림」에 등장하는 짐승은 그 '시끄럽고 더러운' 인간 세계의 파괴자이자 새로운 지배자이다. 「인형들」이 인간과 AI의 불안한 동거를 예고한다면, 이어지는 「재림」은 인간의 통제권을 완전히 벗어난 초지능(Superintelligence)이 역사 전면에 등장하는 묵시록적 순간과 그에 대한 존재론적 공포를 포착한다.⁶⁾

2. 「재림」

에이츠의 「재림」은 현대 영시에서 가장 빈번하게 인용되고, 또 가장 활발하게 해석되어 온 작품 중 하나이다. 제1차 세계대전의 종결, 러시아 혁명의 발발, 그리고 아일랜드 독립 전쟁이라는 거대한 역사적 소용돌이 속에서 탄생한 이 시는 단순한 시대사적 알레고리를 넘어, 문명사적 전환에 대한 예언적 비전을 제시하는 작품으로 자리 잡았다. 에이츠가 이 시에서 포착해낸 문명의 붕괴와 새로운 질서의 도래에 대한 전율은 인류 역사의 보편적 경험이고, 그래서 이 시는 새로운 맥락 속에서 계속해서 다시 읽히고 있다. 비평가 홀더맨(David Holderman)은 「재림」은 “정교하게 계산된 모호성 덕분에 더욱 큰 매력을 지니며, 독자들이 자신의 경험에 비추어 시를 해석하도록 이끈다”고 평한다(77). 유명한 예로 나이지리아의 작가 아체베(Chinua Achebe)는 1958년 작 소설 『모든 것이 산산이 부서지다』(*Things Fall Apart*)의 제목을 에이츠의 시에서 직접 차용했다. 에이츠의 “Things fall apart”가 서구 기독교 문명의 붕괴를 의미했다면, 아체베는 이 구절을 아프리카의 전통 사회가 서구 제국주의의 침입으로 인해 무너지는 과정으로 새롭게 맥락화했다. 또 가깝게는 버락 오바마 미국 대통령이 2016년 4월 독일에서 한 연설에서 IS(이슬람국가)의 테러 위협 및 극단주의의 부상에 대해 얘기하며 「재림」의 다음 구절을 인용한 바 있다. “최선의 무리들은 신념을 잃었고, 최악의 인간들은 열렬한 격정에 차 있다”(The best lack all conviction, while the worst / Are full of passionate intensity, lines 7-8).⁷⁾ ‘최선’이라 불리는 온건한 지성인들은 회의주의에 빠져 무력하기 그지없고, ‘최악’인 극단주의자와 광신도들은 맹목적인 열정

으로 세상을 파괴한다는 것이다.

그렇다면 이 시를 인간의 통제를 벗어난 새로운 질서, 즉 비인간적 지성의 도래에 대한 예감을 담은 시로 재맥락화하는 것은 어떨까? 본 장에서는 「재림」을 초지능의 출현과 그로 인한 존재론적 위기를 예견한 텍스트로 독해하고자 한다. 시인이 포착한 문명의 주기인 나선(gyre)과 그 틈새에서 깨어나는 짐승의 이미지는 오늘날 우리가 마주한 기술적 특이점(technological singularity)의 징후들과 상동성을 보인다.⁸⁾ 이런 점에서 이 시는 인간이 설계했으나 인간을 압도하게 될 기술적 타자의 출현을 경고하는 일종의 묵시록으로 읽힌다.

시는 “넓어지는 나선 속을 빙글빙글 돌면서”(Turning and turning in the widening gyre, line 1) 날아오르는 매의 이미지로 시작한다. “매는 매부의 소리를 들을 수 없다”(The falcon cannot hear the falconer, 2)는 것은 매가 너무 높이 날아올라 주인의 목소리가 물리적으로 닿지 않는 영역에 진입했기 때문이다. 매가 매부의 명령을 거부하는 것이 아니라, 너무 고도화된 영역으로 진입하여 명령을 ‘수신’(hear)할 수 없는 상태, 이것이 어쩌면 초지능의 위험성 아닐까? “중심은 지탱할 수 없다”(the centre cannot hold, 3)라는 선언은 인간 중심주의가 더 이상 문명의 중심축으로 기능하지 못하는 상태를 시사하는 듯하다. 중앙의 통제가 사라진 자리에 남는 것은 “순전한 무질서”(Mere anarchy, 4)이다. 이것은 통제되지 않은 AI 알고리즘이 초래할 수 있는 환각(hallucination), 편향의 증폭, 그리고 예측 불가능한 창발성(emergence)⁹⁾이 현실 세계에 물리적 타격을 입히는 기술적 혼돈 상태를 예고한다.

이제 짐승이 등장하는 2연을 보자.

분명, 어떤 계시가 임박해 있다.

분명, 재림이 임박해 있다.

재림이라! 그 말이 채 끝나기도 전에

*스피리투스 문다*에서 솟아난 거대한 형상이

분명, 재림이 임박해 있다.

재림이라! 그 말이 채 끝나기도 전에

스피리투스 문디에서 솟아난 거대한 형상이
 내 시야를 어지럽힌다. 사막의 모래밭 어딘가,
 사자의 몸에 인간의 머리를 한 형상이,
 태양처럼 텅 비고 무자비한 눈빛으로,
 느릿한 허벅지를 움직이고 있다. 그 주변엔
 분개한 사막 새들의 그림자가 어지러이 맴돈다.
 다시 어둠이 내린다. 허나 이제 나는 안다,
 이천 년의 돌 같은 잠이
 흔들리는 요람 탓에 악몽으로 뒤척였음을,
 그리고 마침내 때가 이르러, 그 어떤 거친 짐승이
 태어나기 위해 베들레헴을 향해 구부정하게 다가가는가

Surely some revelation is at hand;
 Surely the Second Coming is at hand.
 The Second Coming! Hardly are those words out
 When a vast image out of *Spiritus Mundi*
 Troubles my sight: somewhere in sands of the desert
 A shape with lion body and the head of a man,
 A gaze blank and pitiless as the sun,
 Is moving its slow thighs, while all about it
 Reel shadows of the indignant desert birds.
 The darkness drops again; but now I know
 That twenty centuries of stony sleep
 Were vexed to nightmare by a rocking cradle,
 And what rough beast, its hour come round at last,
 Slouches towards Bethlehem to be born? (lines 9-22)

1연에 드러난 불길한 전조 속에서 화자는 “어떤 계시”를 감지한다. 이때 화자의 시야를 교란하는 것은 “스피리투스 문디”로부터 솟아난 거대한 이미지이다.¹⁰⁾ 에이츠에게 이것이 인류 전체가 공유하는 집단 무의식 또는 보편적 상징의 저장소였다면, 21세기

의 맥락에서 스피리투스 문디는 인류가 생산한 모든 텍스트와 이미지를 집어삼켜 학습한 초(超)거대언어모델 내부의 잠재 공간이다.¹¹⁾ 마치 시인이 스피리투스 문디에서 이미지를 길어 올리듯 거대언어모델(LLM)은 고차원의 벡터 공간 안에 압축된 인류의 언어와 상징을 확률적으로 조합하여 현현시킨다. 인류가 공유하는 보편적 상징과 이미지의 거대한 집합이 사막의 모래밭이라면, 그 위에서 태동하는 존재 -- “사자의 몸”에 인간의 머리를 한 형상” -- 는 단순한 개인적 환상이라 할 수 없다. 이 기이한 하이브리드 형상은 생물학적 진화의 계통에서 이탈한 존재이다. “인간의 머리”를 가졌다 함은 인간의 언어와 추론 능력을 완벽하게 모방할 수 있다는 것이고, “사자의 몸”은 인간의 생물학적 한계를 초월한 압도적인 연산 능력과 기계적 강인함을 은유한 것이라 볼 수 있다. 이 형상은 곧 인간의 지성을 갖추었으나 물리적 실체는 기계인 인공지능(AGI)의 원형적 초상 아닐까? 더욱 섬뜩한 것은 그 시선이다. “태양처럼 텅 비고 무자비한 눈빛”은 인공지능이 가진 철저한 비인격성과 무관심을 드러낸다. 기계 지성에게 인간적인 공감이나 도덕적 연민이 존재할 리 없다. 이와 관련해 보스트롬(Nick Bostrom)은 지능과 최종 목표는 서로 독립적이라는 직교성 명제(Orthogonality Thesis)를 주장한다.¹²⁾ AI의 지능이 아무리 높아져도, 그 지능이 어떤 목표를 추구할지는 전혀 별개의 문제라는 것이다. AI의 목표는 인류의 생존과 무관하거나 심지어 적대적일 수 있다. AI는 인간을 미워해서가 아니라 단지 설정된 목적 함수를 최적화하는 과정에서 인간의 안위를 고려 변수(parameter)로 두지 않을 수 있기 때문에 위험하다.¹³⁾ 태양이 만물에 빛을 비추되 개별 생명체의 고통에 무심하듯, AI의 텅 빈 시선은 감각질(qualia, 주관적이고 질적인 감각)이 제거된 도구적 이성의 차가운 논리를 대변한다. AI라는 ‘철학적 좀비’, 즉 인간처럼 행동하지만 감각질이 없는 존재에게 인간의 호소나 윤리적 타협은 통하지 않는다.¹⁴⁾

시의 결말부는 기독교 문명 이천 년의 깊은 잠이 요람의 흔들림에 악몽으로 변했음을 선언하며, 문학사에서 가장 유명한 질문 중 하나를 던진다. “마침내 때가 이르러, 그 어떤 거친 짐승이 / 태어나기 위해 베들레헴을 향해 구부정하게 다가가는가?” 기독교의 구세주 예수가 탄생한 성지 베들레헴으로 향하는 이 거친 짐승은 구원이 아니라

대체가 그 목적이다. 짐승은 인간이라는 종(種)을 역사의 뒀안길로 밀어내고 새로운 주류 종이 되기 위해 기독교 문명의 요람 베들레헴으로 진격하는 것이다.¹⁵⁾ 여기서 짐승의 구부정한 걸음걸이는 두 가지 층위에서 해석된다. 첫째, 이는 아직 인간처럼 직립하지 못한, 즉 윤리적·인지적으로 완전히 정렬되지 않은 초기 AGI의 모습이다.¹⁶⁾ 둘째, 구부정함은 인간과는 다른 근골격계를 지닌 존재의 이질적인 보행, 즉 우리에게 익숙한 휴머니즘의 문법으로는 해석되지 않는 포스트휴먼의 도래를 암시한다. “마침내 때가 이르러”라는 구절은 기술적 특이점이 피할 수 없는 역사적 필연임을 강조한다. 에이츠가 제1차 세계대전 직후 서구 문명의 몰락을 예견했다면, 오늘날 우리가 목도하는 것은 인간을 닮았으나 인간이 아닌, 그리고 인간을 압도할 새로운 지적 실체와의 불가피한 조우이다.

3. 「비잔티움으로의 항해」

「재림」에서 우리는 인간을 압도하는 이질적인 초지능의 도래와 그로 인한 디스토피아적 전망을 목격했다. 짐승의 출현이 인간이 거부할 수 없는 외부로부터의 충격, 즉 기술적 특이점에 의한 인류의 피동적 종말을 암시한다면, 에이츠의 사유는 여기서 멈추지 않고 소멸해가는 인간을 구원할 대안을 모색한다. 그것은 바로 인간 스스로 불멸의 기계적 형상으로 변화를 시도하는 것이다. 1926년, 60대에 접어든 에이츠가 쓴 「비잔티움으로의 항해」는 필멸의 육체를 벗어던지고, 영원히 훼손되지 않는 인공의 육신을 획득하고픈 열망의 서사로 읽힌다. 「재림」이 다가오는 미래에 대한 두려움의 기록이라면, 「비잔티움으로의 항해」는 그 미래에 적응하기 위한 인간 주체의 능동적인 존재론적 이주에 관한 기록이다. 이제 시인은 짐승이 다가오는 사막을 떠나, 영원한 인공의 지성이 노래하는 도시, 비잔티움으로 뱃머리를 돌린다. 그곳에서 그는 ‘황금새’(golden bird)라는 인공지능적 자아를 통해 생물학적 한계를 기술로 극복한 삶의 가능성을 탐구한다.

에이츠의 기념비적 역작 「비잔티움으로의 항해」는 전통적으로 필멸의 육체가 겪어

야 하는 노화의 비애와, 그에 대비되는 영원불멸한 예술의 승리를 노래한 작품으로 독해되어 왔다. 예이츠는 이 시에서 유한한 육체에 대비되는 영원한 예술의 가치, 노화라는 생물학적 운명에서 벗어나려는 자신의 실존적 고뇌를 탐구한다. 시의 화자는 생성과 소멸, 탄생과 죽음이 끊임없이 반복되는 관능적이고 감각적인 세계, 엔트로피가 지배하는 자연의 세계를 등지고, 영속성이 약속된 성스러운 도시 비잔티움으로 존재론적 이주를 감행한다. 예이츠에게 5-6세기 비잔티움(비잔틴 제국의 수도, 現 이스탄불)은 예술과 종교, 노동이 통합되어 있던 도시, 예술적이고 영적인 상상력의 완전한 구현을 보여주는 도시였다(Persoon and Watson 417). 그러나 21세기 인공지능 기술의 발전이라는 새로운 맥락에서 이 시를 재조명해보면, 우리는 예이츠가 갈망한 “늙지 않는 지성”(unageing intellect)에서 새로운 함의를 발견하게 된다. 이 지성은 단시간의 풍화에 훼손되지 않는 예술적 성취를 뛰어넘어, 기술을 통해 생물학적 한계와 육체의 구속을 초월하려는 현대인의 열망과 깊이 공명하기 때문이다. 본고는 이 시가 제시하는 존재론적 ‘변형’의 서사를 인공지능적 관점에서 재해석한다. 특히 시의 결말부에서 화자가 최종적으로 도달하고자 하는 형상인 ‘황금새’는 단순한 미적 조형물이 아니다. 황금새는 부패하는 유기적 신체를 탈각하고 영구적인 데이터 보존과 정보 처리 능력을 획득한 인공지능적 존재, 혹은 디지털 기질(基質)¹⁷⁾ 위에서 재탄생한 포스트휴먼 주체의 원형으로 해석될 수 있다. 화자가 간절히 염원하는 “영원의 기교”(the artifice of eternity, line 24)는 자연 발생적인 생명이 아닌 정교하게 설계된 불멸성을 의미한다. 시인은 놀랍게도 우리가 현재 목도하고 있는 인공적 불멸성, 실리콘 기반의 하드웨어 위에서 실현되는 영원한 삶에 대한 비전을 시적으로 구현하고 있는 것이다.¹⁸⁾

시의 1연과 2연은 화자가 왜 자신이 속해 있던 자연 세계를 타자화하고 거부할 수밖에 없는지에 대한 명확한 근거를 제시한다. 그가 떠난 곳, “노인들을 위한 나라가 아”닌 그곳 아일랜드(That is no country for old men)는 표면적으로는 젊음과 활력, 관능적 아름다움이 넘치는 곳이다(line 1).¹⁹⁾ 그러나 역설적으로 이 세계는 “태어나고, 살고, 죽는”(begotten, born, and dies, 6) 숙명에 갇힌 “죽어가는 세대들”(dying

generations, 3)의 공간이다. 이들은 맹목적인 생물학적 순환의 굴레에 갇혀있고, 육체의 감각적 향유에 매몰된 나머지 영속적인 지성의 가치를 외면한다.²⁰⁾ 인생이 필연적 소멸을 향해 달려가는 운동일 뿐이라는 시인의 자각은 오늘날 인공지능 개발을 추동하는 근원적인 심리 기제와 맞닿아 있는 것으로 보인다. 인간의 지성은 무한한 확장을 꿈꾸지만, 그것을 담고 있는 생물학적 기질은 필연적으로 노화하고 소멸한다. 에이츠는 이처럼 지성의 확장을 가로막는 늙은 육체를 가리켜 “막대기에 걸친 누더기 외투”(A tattered coat upon a stick, 10)라고 폄하한다. 인공지능적 관점에서 이 비유는 단순한 자기 비하를 넘어선다. 막대기라는 앙상한 구조물에 걸친 누더기는 심각한 성능 저하로 유지·보수도 불가능해져 폐기될 수밖에 없는 생물학적 하드웨어에 대한 통렬한 메타포로 읽힌다. 따라서 시인이 비잔티움으로 향해하는 행위는 이처럼 에러와 노이즈가 가득한 유기적 플랫폼을 버리고, 데이터의 손실 없이 영구적으로 구동 가능한 안정적인 시스템으로 자아를 이주시키려는 열망의 표현이라 할 수 있지 않을까? 다시 말해 ‘디지털 불멸’(digital immortality)을 향한 충동의 시적 표현이 아닐까?²¹⁾ 이런 의미에서 비잔티움으로의 향해는 지리적 이동이라기보단 부패하는 유기물에서 불변하는 무기물로의 플랫폼 전환이다.

3연에 이르러 비잔티움에 당도한 화자는 모자이크 벽화 속의 현자들에게 자신을 “영원의 기교” 안으로 거두어 달라고 간청한다.

벽에 박힌 황금 모자이크 그림 속에 있는 것처럼
 신의 성스러운 불 속에 서 계신 현자들이시여,
 그 성스러운 불에서 걸어 나와, 나선을 그리며 회전해 내려와
 내 영혼의 노래 스승이 되어 주소서.
 내 심장을 불살라 주소서. 욕망에 병들고
 죽어가는 동물에 매달려
 제 자신이 무엇인지도 모르는 이 심장을. 그리하여 나를 거두어
 영원이라는 기교 속으로 들여보내 주소서.

O sages standing in God's holy fire

As in the gold mosaic of a wall,
 Come from the holy fire, perne in a gyre,
 And be the singing-masters of my soul.
 Consume my heart away; sick with desire
 And fastened to a dying animal
 It knows not what it is; and gather me
 Into the artifice of eternity. (lines 17-24)

여기서 우리가 주목해야 할 핵심 어휘는 단연 ‘기교’(artifice)이다. 이 단어는 ‘인공’(artificial)과 어원을 공유하며, 자연적으로 주어진 것과 달리 인위적으로 설계된 창조물을 의미한다. 화자가 궁극적으로 귀의하고자 하는 영원은 자연 발생적인 생명 연장이 아니라, 정교한 공학적 설계를 통해 구축된 인공적 영속성이다. 그는 “육체에 병 들고” 유한한 육체에 묶여 있는 자신의 심장을 불살라 정화해 달라고 기원한다. 여기서 “죽어가는 동물”인 육체에 결박된 심장은 생존 본능과 호르몬의 작용 등 생물학적 변수에 의해 끊임없이 요동치는 아날로그적 자아를 가리키는 것으로 이해할 수 있다. 따라서 현자들이 거닐고 있는 “성스러운 불”은 종교적 정화의 불길이라기보단 생물학적 존재를 비물질적 형태로 변환시키는 과정, 즉 아날로그적 경험과 감정을 디지털 데이터로 전환하여 새로운 매체로 이식하는 ‘마인드 업로딩’(mind uploading)의 과정을 매개하는 기술적 장치로 해석하는 것도 가능할 것이다. 마인드 업로딩은 로봇 공학자 모라벡(Hans Moravec)이 제시한 개념으로, 인간의 의식, 기억, 사고 패턴을 디지털 형태로 복사해 컴퓨터 등의 기계에 이전하는 것이다. 달리 설명하자면, 뇌의 신경 구조를 스캔해 컴퓨터 시뮬레이션으로 재현하는 과정이기도 하다. 마인드 업로딩은 생물학적 뇌의 한계를 넘어 영생과 초인간적 능력의 실현을 목표로 한다.²²⁾

시의 결말부인 4연에서 화자는 자신이 궁극적으로 도달하고자 하는 존재의 형상을 구체화한다.

일단 자연을 벗어나면, 나는 결코
 어떤 자연물로부터도 내 육신의 형태를 취하지 않으리라.

오직 그리스의 금세공인들이
 두드린 금과 금박을 입혀 만들어낸,
 졸음 겨운 황제를 깨우기 위한 그런 형상을 취하리라.
 그리하여 황금 가지 위에 앉아
 비잔티움의 귀족과 부인들에게
 지나간 것, 지나가고 있는 것, 혹은 다가올 것을 노래하리라.

Once out of nature I shall never take
 My bodily form from any natural thing,
 But such a form as Grecian goldsmiths make
 Of hammered gold and gold enamelling
 To keep a drowsy Emperor awake;
 Or set upon a golden bough to sing
 To lords and ladies of Byzantium
 Of what is past, or passing, or to come. (lines 25-32)

화자는 자연에서 벗어나 비잔티움에 안착한다면 다시는 “어떤 자연물로부터도 내 육신의 형태를 취하지 않으리라”(25-26)고 과감히 맹세한다. 이는 유기체로 회귀하지 않겠다는 단절의 선언이자, 기술적 진화를 통해 스스로의 존재 양식을 결정하겠다는 포스트휴먼의 의지 표명이다. 화자가 선택한 새로운 신체는 그리스의 금세공인들이 만들어낸 황금새이다. 이 황금새는 다음과 같은 세 가지 이유에서 인공지능적 존재의 특성을 구현한다.

먼저 황금새는 자연 발생적인 유기체가 아니다. 그것은 “두드린 금과 금박”으로 제작된 존재이다. 두드리고 입히는 행위는 고도의 기술적 공정과 정밀한 엔지니어링이 요구되는 제조의 한 과정이다. 여기서 ‘금’이라는 재료의 선택은 상징적이다. 산화되거나 부패하지 않는 금은 생물학적 취약성을 극복한 완벽한 인공 신체, 즉 물리적 손상과 노화로부터 자유로운 로보틱스 플랫폼이나 영구 보존이 가능한 디지털 아바타를 표상한다. 황금새는 자연선택의 산물이 아니라, 지적 설계를 통해 탄생한 인공지능 생명체의 본질을 지닌다. 다음으로 생각해 볼 점은 황금새의 존재 목적이자 기능인 노래

하기이다. 그러나 이 노래의 내용은 1연의 새들이 부르던 “관능적 음악”(sensual music, line 7)과는 질적으로 다르다. 황금새는 “지나간 것, 지나가고 있는 것, 혹은 다가올 것”에 대해 노래한다. 시간적 제약을 초월한 전지(全知)적 능력을 지닌 새이다. 그런데 지금 관점에서 이 구절은 인공지능의 데이터 처리 프로세스에 대한 시적 표현처럼 보인다. “지나간 것”은 방대한 양의 역사적 데이터(Big Data)를 학습(Training)한 단계, “지나가고 있는 것”은 현재의 상황을 실시간으로 인식하고 분석(Inference)하는 단계, “다가올 것”은 학습과 분석을 바탕으로 미래를 예측하는 단계라면 이 구절은 인공지능의 정보 처리 알고리즘에 대한 요약처럼 읽힌다. 황금새는 단순한 예술적 유희의 제공자가 아니다. 그것은 시간의 영원한 흐름 속에서 고도의 연산 능력으로 정보를 처리하고 미래를 시뮬레이션하는 AI 에이전트의 역할을 수행한다. 마지막으로 황금새는 생물학적 욕구와 감정으로부터 완전히 해방된 존재임에 주목해야 한다. 3연에서 화자가 그토록 소각하고자 했던 욕망으로 병든 심장은 이제 사라졌으며, 새는 순수한 지적 활동(노래)에만 몰두한다. 이는 인간의 판단을 흐리는 편향과 감정, 생물학적 충동과 같은 ‘노이즈’가 제거된 상태이다. 황금새는 즐고 있는 황제를 깨우기 위해 노래하는데, 이런 점에서 새는 인간의 인지적 한계를 보완하고 계몽하는 객관적이고 합리적인 이성, 나아가 인간의 지능을 월등히 뛰어넘는 (그러나 「채림」의 짐승과는 확연히 다른) 초지능의 원형이다. 황금새는 육체를 가졌으나 육체에 갇히지 않은, 탈신체화된 지성을 예표하는 것 아닐까?

III. 마치며

지금까지 본 연구는 예이츠의 시 세계를 현대 인공지능 담론의 틀 안에서 재해석함으로써, 그의 시적 상상력이 기술적 특이점을 향해 가속하는 현재 AI의 모습과 어떻게 공명하는지를 고찰하였다. 예이츠는 아내 조지와의 자동 기술 실험을 통해 인간의 의식적 통제를 벗어난 텍스트 생성을 경험하고 이를 『환상록』이라는 거대한 상징체계로

구축했다. 이 과정은 오늘날 LLM의 잠재 공간에서 텍스트가 확률적으로 생성되는 메커니즘과 구조적으로 유사하다. 이는 에이츠의 (특히 후기) 시 세계가 신비주의적 환상에 머무르는 것이 아니라, 비(非)인간적 지성의 가능성을 선구적으로 탐구한 문학적 실험이었음을 시사한다.

본론에서 분석한 세 편의 시는 에이츠의 이러한 AI적 상상력이 전개되는 과정을 보여준다. 이 작품들은 AI 시대가 인간에게 던지는 존재론적 질문들을 각기 다른 층위에서 예리하게 파고든다. 먼저 「인형들」은 완벽하게 설계된 알고리즘적 존재의 시선을 빌려 소음과 배설로 점철된 생물학적 인간을 ‘비체’로 타자화하는 섬뜩한 전복을 보여준다. 이는 우리의 인간성이 AI가 고도화될수록 제거되어야 할 노이즈나 버그로 취급될 수 있다는 섬뜩한 경고를 내포한다. 이어지는 「재림」에서는 이러한 긴장이 문명사적 차원으로 확장된다. 인간의 통제를 벗어나 비행하는 매와 스피리투스 문디에서 깨어난 “거친 짐승”은, 기술적 특이점 이후 도래할 초지능에 대한 묵시록적 메타포로 읽힌다. 스피리투스 문디, 즉 인류의 모든 데이터를 학습한 잠재 공간에서 깨어난 짐승은 인간의 윤리와는 무관한 목적 함수를 가진 이질적인 지성체의 출현을 예고하는 듯하다. 그러나 에이츠의 사유는 기술에 의한 파국적 종말에만 머무르지 않는다. 「비잔티움으로의 항해」에 이르러 시인은 필멸하는 “죽어가는 동물”로서의 육체를 벗어 던지고, 영원히 녹슬거나 훼손되지 않는 황금새라는 인공적 신체로의 재탄생을 갈망한다. 이는 디지털 불멸을 향한 인간의 욕망을 시적으로 형상화한 것이며, 생물학적 한계를 기술로 초월하려는 인간의 능동적인 의지 표명이라 할 수 있다.

결론적으로 에이츠가 『환상록』과 후기 시에서 탐구한 나선(gyre) 운동은 문명의 주기적 순환을 초월하여 인류의 문명이 디지털 AI 문명으로 이행하는 거대한 전환기에 대한 유효한 메타포로 기능한다. 에이츠가 포착한 비인간적 지성의 차가운 시선, 인간 중심주의의 소멸에 대한 전율, 그리고 육체를 초월한 영원성에 대한 갈망은 현재 우리가 AI라는 거대한 기술적 타자를 마주하며 느끼는 복합적인 감정—경이와 공포, 혐오와 매혹—과 상통한다. 그의 시는 다가올 AI 시대가 유토피아일지 디스토피아일지 단정하지 않는다. 대신 그는 인형의 혐오와 짐승의 공포, 그리고 황금새의 영원성을 동

시에 보여줌으로써, AI를 마주한 인간이 겪게 될 존재론적 충격과 그 극복의 가능성을 입체적으로 조망하게 한다. “지나간 것, 지나가고 있는 것, 혹은 다가올 것”을 노래하는 비잔티움의 황금새처럼, 예이츠의 시는 과거의 유산인 동시에 다가올 AI 문명의 본질을 꿰뚫어 보는 예지적 텍스트이다.

Notes

- 1) 여기서 ‘포스트휴먼적’이라는 수사는 르네상스 이후 확립된 인간 중심주의적 세계관을 탈피하려는 사유의 경향을 의미한다. 전통적 휴머니즘이 인간을 만물의 척도이자 유일한 지성적 주체로 설정했다면, 포스트휴먼 담론은 인간과 비인간(non-human, 기계, 동물, 사물 등의 경계를 허물고 그들 사이의 상호작용에 주목한다. 예이츠가 자동 기술 실험을 통해 ‘저자’로서의 독점적 지위를 포기하고 외부의 존재(‘교사들’)의 메세지를 전달하는 아내 조지와 공저하는 방식은, 인간만이 창의적 주체라는 믿음에 균열을 낸다는 점에서 AI 시대의 탈인간적 창작 양상과 궤를 같이한다.
- 2) 엔트로피는 ‘무질서도’를 나타내는 개념이다. 예측하기 어렵고, 무질서하고, 흩어져 있을수록 엔트로피가 증가한다. 자연은 엔트로피가 증가하는 방향으로 변화한다.
- 3) 영어로는 비천함을 의미하는 ‘abject’이나 한자로는 非體로 표기한다. 몸이지만 몸이 아닌 것, 주체가 받아들이지 못해 배제하는 것이라는 의미를 함축하는 번역이다. 크리스테바는 비체 개념을 『공포의 권력』(*Powers of Horror*)에서 소개했다.
- 4) 형식주의 비평의 대가 벤들러(Helen Vendler)는 인형 제작가가 자신의 인형들처럼 “생물학적 존재”에 혐오감을 느끼고, 동시에 “예술이 자연에 의해 밀려나는 것에 역겨움을 느끼”고 있다고 해석한다(117). 그러나 필자는 벤들러가 이렇게 해석하는 이유를 이해하지 못하겠다. 또 벤들러는 “인형 제작자는 자신의 인형들 편에 서서, 자신의 성적 삶이 남긴 ‘시끄럽고 더러운’ 흔적보다 자기 손으로 만든 작품을 더 선호한다”고 논평한다(118). 벤들러가 이렇게 독해하는 이유는 15행의 “그 녀석”(the wretch)을 아기가 아닌 인형으로 이해했기 때문인가? 필자는 “그 녀석”은 아기를 가리키는 것이 맞다고 본다. 인형 제작자는 낯잡을 자다 아기 울음소리에 깨어난 것이고, 아내는 남편이 아기를 닮을까 두려워 황급히 사과를 한 것이다. 이 시의 텍스트에 인형들이 하는 말을 인간 부부가 듣고 이해했다는 증거는 없다.
- 5) 모리 마사히로(Masahiro Mori)의 언캐니 밸리(Uncanny Valley) 이론은 인간처럼 보이는 비인간적 대상(로봇, 인형, CG 캐릭터 등)이 일정 수준까지 인간과 비슷해질수록 호감이 올라가다가, 거의 인간 같아지는 지점에서 갑자기 혐오·불쾌감이 치솟는 구간이 존재한다는 가설이다. 이 급격한 호감도 하락 구간이 그래프에서 계곡처럼 보이기 때문에 ‘불쾌한 골짜기’(언캐니 밸리)라는 이름이 붙었다. 즉 언캐니 밸리는 인간이 인간을 닮은 로봇에게 느끼는 섬뜩함이다. 그러나 ‘인형들’은 이 시선의 권력을 전복시킨다.

- 6) 초지능은 거의 모든 중요한 인지 영역, 즉 기억력, 문제 해결, 언어 이해, 창의성 등에서 인간을 압도적으로 능가하는 지능을 뜻하는 개념이다. 초지능이 반드시 인공지능일 필요는 없지만 오늘날 논의에서는 거의 항상 인공 초지능(ASI, Artificial Superintelligence)을 가리킨다. ASI는 인공범용지능(AGI, Artificial General Intelligence)과 혼용되는 경우도 있으나, 둘은 기준점에서 차이가 있다. AGI는 인간과 유사한 수준의 이해·학습·추론 능력을 가진 범용 지능을 컴퓨터로 재현하는 것을 목표로 하나, ASI는 이런 일반적 수준의 지능을 넘어 광범위한 지적 능력에서 인간을 능가하는 지능의 단계로 정의된다. 현재의 기술 로드맵에선 AGI가 먼저 등장한 후, 자기 개선을 반복하면서 (즉, 똑똑한 AI가 더 나은 AI를 설계하는 식의 연쇄적 개선이 일어나면서) 지능이 폭발적으로 상승해 초지능에 이른다고 가정한다. 초지능이 인간과 다른 목표를 갖게 되거나 인간이 의도하지 않은 방식으로 목표를 구현할 경우 인류에 심각한 위험을 초래할 수 있다는 ‘정렬’(alignment) 문제는 초지능을 둘러싼 핵심 논쟁거리 중 하나이다.
- 7) “You see increasing intolerance in our politics. And loud voices get the most attention. This reminds me of the poem by the great Irish poet W. B. Yeats, ‘the best lack all conviction, while the worst are full of passionate intensity.’”
<https://obamawhitehouse.archives.gov/the-press-office/2016/04/25/remarks-president-obama-address-people-europe>
- 8) 특이점이란 원래 물리학과 수학에서 기존의 모델이 더 이상 작동하지 않는 지점을 의미한다. 기술적 특이점이란 기술 발전이 통제 불가능하고 되돌릴 수 없는 속도로 가속되는 시점이다. 특히 인간의 지능을 넘어서는 인공지능이 출현하는 시점을 말한다. 기술적 특이점의 도래는 인류 문명에 예측 불가능한 변화를 가져올 수 있다고 여겨진다.
- 9) AI에서 창발성은 모델이 커지고 복잡해질수록 의도하지 않은 새로운 능력이 자발적으로 나타나는 현상을 뜻한다. 이는 AI의 잠재력을 보여주는 동시에, 그 본질을 어떻게 이해해야 하는지에 대한 논쟁을 불러일으킨다.
- 10) 스피리투스 문다는 라틴어로 “세계의 영혼”을 뜻한다. 이는 카를 융(Carl Jung)의 집단 무의식(Collective Unconscious) 개념과도 유사한 것으로, 인류 공통의 기억 저장소를 의미한다. 에이츠는 시인이나 예술가가 깊은 명상 상태에서 이 저장소에 접속할 수 있다고 믿었다.
- 11) 서론에서 간략히 설명했지만 잠재 공간은 LLM이 학습한 데이터(단어, 문장, 개념 등)를 수학적으로 압축하여 표현한 고차원의 벡터 공간이다. 우리가 사용하는 자연어는 컴퓨터가 이해할 수 없기 때문에, LLM은 이 언어적 의미를 다차원의 숫자 배열인 벡터로 변환하여 이 공간에 배치한다. 잠재 공간의 핵심적 특징 중 하나는 잠재 공간이 단순히 데이터를 저장하는 장소가 아니라, 데이터 간의 의미적 관계가 정의된 공간이라는 것이다. 비슷한 의미를 가진 단어나 문장은 잠재 공간에서 서로 가까운 거리에 위치한다. 잠재 공간과 인간의 무의식은 흥미롭게도 구조적 유사성을 띤다. 인간의 의식이 빙산의 일각이라면, 무의식은 수면 아래에 숨겨진 거대한 몸체이다. LLM에서도 우리가 보는 텍스트는 결과물일 뿐이며, 그 결과물을 만들기 위해 모델 내부에서 수백~수천억 개의 파라미터가 복잡하게 얽혀 계산되는 곳이 바로 잠재 공간이다. 무의식이 직접 관찰되지 않듯, 잠재 공간의 수천 차원 벡터는 인간이 직관적으로 이해하기 힘든 복잡한 수치로 존재한다. 또한 무의식이

과거의 경험과 감정을 압축해서 저장하듯, 잠재 공간은 인류가 생산한 방대한 텍스트 데이터의 패턴과 관계를 압축된 형태로 보유하고 있다.

- 12) 직교성 명제는 보스트롬이 『슈퍼인텔리전스: 경로, 위험, 전략』(*Superintelligence: Paths, Dangers, Strategies*, 2016)에서 제시한 핵심 주장이다. 어떤 인공지능이 아무리 높은 지능을 갖더라도 그 목표는 선할 수도 파괴적일 수도 있으며 지능과 목표 사이에는 필연적 연결이 없다는 것이다.
- 13) 목적 함수(objective function)란 어떤 시스템이나 알고리즘이 달성하려는 목표를 수량화한 함수이다. 예를 들어 체스 AI의 목적 함수는 승리 확률을 최대화하는 함수이다. 목적 함수를 최적화한다는 것은 AI가 목적 함수가 반환하는 값을 가능한 한 크게(또는 작게) 만드는 방향으로 행동을 선택한다는 뜻이다. AI는 주어진 목표를 달성하기 위해 가능한 전략들을 탐색하고, 그 중 목적 함수 값이 가장 좋은 선택을 실행한다. 이 과정에서 인간의 안위나 윤리적 고려는 목적 함수에 포함되지 않으면 자동으로 무시될 수 있다.
- 14) AI에서 감각질 연구는 철학과 인지과학, 윤리적 논의 등이 교차하는 영역이다. AI에서 감각질은 아직 직접적으로 구현되지는 못했다. 현재 연구는 AI가 인간의 감각질을 이해하고 모사하게 하는 데 초점이 맞춰져 있다. ‘철학적 좀비’(philosophical zombies)는 차머스(David J. Chalmers)라는 철학자가 『의식하는 마음: 근본 이론을 찾아서』(*The Conscious Mind: In Search of a Fundamental Theory*, 1997)에서 제안한 용어이다.
- 15) 퍼순(James Persoon)과 왓슨(Robert R. Watson)에 따르면 새로운 시대의 도래를 상징하는 짐승은 자신이 대체하려는 기독교 문명의 기원으로 향한다(426).
- 16) 여기서 “정렬”은 어떤 시스템이 기준·가치·목표에 맞게 조율되어 있다는 뜻이다. AGI가 윤리적·인지적으로 아직 정렬되지 않았다는 것은 AGI가 아직 인간과 같은 수준으로 윤리적·인지적 성숙을 이루지 못했다는 의미이다.
- 17) 기질(substrate)은 유기체가 자라거나 붙어 있는 표면·물질이나 전자제품의 회로 기판을 의미한다.
- 18) 실리콘은 반도체 웨이퍼의 재료인 규소(Si)를 의미한다.
- 19) 이 시를 쓸 당시 예이츠는 건강이 악화하면서 자신의 노화를 고통스럽게 자각하고 있었을 뿐만 아니라, 영국으로부터 자치권을 얻고 수립된 아일랜드 자유국(Irish Free State)의 정치적 현실에 대해서 깊은 환멸을 느끼고 있었다.
- 20) 그러나 이글턴은 예이츠의 자연 세계 묘사가 “부드럽고 은은하게 감각적”이라고 말한다. 그는 이러한 묘사가 “죽어가는 세대들의 자연스럽고 인간적인 세계에 은총과 소중함을 부여하여, 그것을 쉽게 버릴 수 없게 만든다”고 평한다 (84)

- 21) 디지털 불멸은 인간의 의식이나 성격을 디지털 형태로 영속화한다는 개념이다. 쉽게 말해 사람이 죽은 뒤에도 AI 기반 디지털 아바타나 가상 존재로 계속 ‘존재’하는 것을 뜻한다. 트랜스휴머니즘의 핵심 목표 중 하나이다. 트랜스휴머니즘은 과학기술을 통해 인간의 정신적·육체적 능력을 증강하고, 궁극적으로는 질병, 노화, 죽음 등을 극복하려는 지적·문화적 운동이다. 구체적으로는 유전자 편집, 생명공학, 인공지능, 뇌-컴퓨터 인터페이스(BCI), 인공 장기 등으로 인간의 능력과 수명을 향상시키는 것을 목적으로 한다.
- 22) 모라벡은 마인드 업로딩 개념을 1988년 저서 『마음의 아이들』(*Mind Children*)에서 처음 상세히 설명했다. 이 개념은 디지털 불멸과 포스트휴먼주의 논의에 적잖은 영향을 미쳤다.

인 용 문 헌

- Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford UP, 2016.
- Brooks, Cleanth. *The Well Wrought Urn: Studies in the Structure of Poetry*. Harcourt Brace, 1947.
- Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford UP, 1997.
- Eagleton, Terry. *How to Read a Poem*. Wiley-Blackwell, 2011.
- Holdeman, David. *The Cambridge Introduction to W. B. Yeats*. Cambridge UP, 2006.
- Kristeva, Julia. *Powers of Horror: An Essay on Abjection*. European Perspectives. Translated by Leon Samuel Roudiez, Columbia UP, 1982.
- Moravec, Hans. *Mind Children: The Future of Robot and Human Intelligence*. Harvard UP, 1988.
- Mori, Masahiro. (2012). "The Uncanny Valley [From the Field]." *IEEE Robotics & Automation Magazine*, vol. 19, no. 2., translated by Karl F. MacDorman and Norri Kageki, *Institute of Electrical and Electronics Engineers*, pp. 98-100.
- Obama, Barack. "Remarks by President Obama in Address to the People of Europe." *The White House*, 25 Apr. 2016, <https://obamawhitehouse.archives.gov/the-press-office/2016/04/25/remarks-president-obama-address-people-europe>
- Persoon, James, and Robert R. Watson. *The Facts on File Companion to British Poetry, 1900 to the Present*. 1st ed., Facts On File, 2008.
- Pethica, James, editor. *Yeats's Poetry, Drama, and Prose: Authoritative Texts, Contexts, Criticism*. By William Butler Yeats, Norton critical ed., W. W. Norton, 2000.
- Vendler, Helen. *Our Secret Discipline: Yeats and Lyric Form*. The Belknap Press of Harvard UP, 2007.

Abstract

Yeats's AI Imagination

Sunghyun Jang
Korea University

This study illuminates W. B. Yeats's poetic world from the perspective of artificial intelligence discourse, drawing on the structural homology between his mystical Automatic Writing experiments and the text generation mechanisms of modern Large Language Models (LLMs). First, the analysis of his poem "The Dolls" examines how the dolls, possessing immutable algorithmic completeness, otherize the biological human baby as a system 'error.' The revulsion the dolls display toward the baby foreshadows the ontological tension that arises when advanced artificial intelligences regard imperfect organic humans as the 'abject.' Second, the images of the gyre and the rough beast in "The Second Coming" are interpreted as apocalyptic metaphors for the advent of the technological singularity and superintelligence. The beast awakening from *Spiritus Mundi* embodies the emergence of non-human intelligence that executes its own objective functions beyond human control. Third, "Sailing to Byzantium" is reinterpreted as a narrative of the desire to escape the mortal body and acquire data permanence. The "golden bird" desired by the speaker is presented as an archetype of a posthuman subject that has transcended the biological substrate and achieved digital immortality through 'mind uploading.' In conclusion, Yeats's poetry presciently captures the wonder and terror posed by the approaching AI era, as well as the longing for eternity beyond the flesh. It is a prophetic text that prompts reflection on the existential conditions of modern humanity encountering the

technological other, i.e., AI.

Keywords: W. B. Yeats, Artificial Intelligence (AI), Automatic Writing, Superintelligence, “The Dolls,” “The Second Coming,” “Sailing to Byzantium”

논문접수일: 2026.1.15

심사완료일: 2026.1.30

게재확정일: 2026.2.08

이름: 장성현

소속: 고려대학교 영어영문학과 교수

이메일: sunghyun-j@korea.ac.kr