

# 드라마<굿 플레이스>속 인공지능 로봇의 상호작용 의례적 해석 Interaction Ritual Interpretation of AI Robot in the TV Show <The good place>

추미선, 유승호  
강원대학교 영상문화학과

Mi-Sun Chu(mschu@kangwon.ac.kr), Seoung-Ho Ryu(shryu@kangwon.ac.kr)

## 요약

인간과 인공지능 로봇과의 관계를 예견하는 쟁점은 강한 인공지능 문제로 많은 전문가들은 인간보다 뛰어난 사고능력을 가진 강한 인공지능이 인간을 정복하는 비극적 결말을 예측하고 있다. 동시에 인공지능 로봇에 대해서 가지고 있는 기대가 드라마나 SF영화와 같은 미디어에 투영하여 나타나게 되면서 인간의 기대에 부응하는 '도덕적으로 선한 인공지능'에 대한 문제 또한 중요한 쟁점이다. 그러나 선한 인공지능에 대한 요구와 완벽한 기술의 구현은 그 문제가 기계에만 국한되지 않고 오히려 인간에게 모든 책임을 지우는 결과로 나타나 인간을 비도덕적인 존재로 몰아가고 인간과 인간의 문제로 전환되어 더 많은 소외와 차별을 낳을 수 있다. 이처럼 기술이 가져오는 결과는 사용하는 인간과 상호작용하며 반응에 따라 그 속성이 결정되고 발전하면서 인간에게 영향을 미친다. 따라서 상호작용을 고려하여 인간의 정서를 고려하는 인공지능 기술도 중요하다. 본 연구는 인간과 인공지능의 관계가 '선한 인공지능'에 대한 요구로 귀결되는 과정과 성공적인 상호작용 과정을 밝히기 위해 이론적으로는 랜달 콜린스의 정서적 에너지로 인간 유대의 형성을 설명하는 상호작용 의례 이론을 통해 설명한다. 방법론으로는 사고실험의 한 유형으로서 드라마 <굿 플레이스>의 재닛과 주변 등장인물들을 통해 선한 인공지능 로봇에 대한 기대가 발생되고 지향되는 지점을 밝히고자 한다.

■ 중심어 : | 상호작용의례 | 랜달콜린스 | 선한인공지능 | 인공지능드라마 | 인공지능로봇 |

## Abstract

The issue of predicting the relationship between humans and AI robots is a 'strong AI' problem. Many experts predict the tragic ending which is a strong AI with superior thinking ability than humans will conquer humans. Due to the expectations of AI robots are projected onto media, the 'morally good AI' that meets human expectations is an important issue. However, the demand for good AI and the realization of perfect technology is not limited to machines. Rather, it appears as a result of putting all responsibility on humans, driving humans into immoral beings and turning them into human and human problems, which is resulting in more alienation and discrimination. As such, the result of technology interacts with the human being used and its properties are determined and developed according to the reaction. This again affects humans. Therefore, AI technology that considers human emotions in consideration of interaction is also important. Therefore, this study will clarify the process that the demand for 'Good AI' in the relationship of AI to humans with Randall Collins' Interaction Ritual Chain. Emotional energy in Interaction Ritual Chain has explained the formation of human bonds. Also, the methodology is a type of thinking experiment and explained through Janet and surrounding characters in the TV show <The good place>.

■ keyword : | Interaction Ritual | Randall Collins | Good AI | AI TV Show | Artificial Intelligence |

\* 이 논문은 2017년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A3A2066149)

접수일자 : 2021년 03월 03일  
수정일자 : 2021년 04월 02일

심사완료일 : 2021년 04월 29일  
교신저자 : 유승호, e-mail : shryu@kangwon.ac.kr

## I. 서론

### 1. 인공지능에 대한 인간의 기대: 선한 반응

미국에서 테슬라의 자율주행 기술은 운전자가 손을 놓고도 가능한 수준이어서 운전 중에 운전자가 숙면을 취하는 것이 큰 문제 되고 있다[1][2]. 이에 따라 자율주행에 대한 법적 규제가 논의되고 있으며 이 문제에 대한 임시방편으로 테슬라의 자동차는 운전자가 운전대를 잡으라는 안전장치를 무시하면 '오토파일럿(autopilot)' 기능을 강제로 종료하도록 설계되어 있다. 그래서 운전자들은 헬퍼(helper) 장치를 사용하여 오토파일럿 기능을 속이려고 한다. 그러자 미 의회에서는 또 다시 백업 드라이버 모니터링을 도입하려고 하고 있다[3]. 속임수를 통해 자율주행 장치를 속이려는 것을 근본적으로 막으려는 제도이다. 기술의 발전은 이렇듯 인간과 기계가 서로 상호작용하며 발전해나간다. 이는 과학기술사회학에서 말하는 기술의 사회적 구성으로 기술 자체로 방향성이 이미 정해져 있기보다 인간이 어떻게 사용하고 받아들이는가에 따라 기술의 방향이 결정되고 발전하며 상호작용한다는 관점이다. 이처럼 기술이 가져오는 결과는 기술을 사용하는 인간의 반응에 따라 속성이 결정되기 때문에 예측하지 못한 방향으로 나타나기도 하며 결과가 크게 달라질 수 있다[4][5].

아마존에서는 공정성을 위해 인공지능을 채용 시스템에 적용하여 성차별을 극복하고자 했다. 그러나 인공지능은 '여성'이 들어간 이력서를 제외하거나 단어 사용에서 남성 기술자들이 자주 쓰이는 동사를 우선적으로 선별하는 등 알고리즘이 공정하게 작동하지 못했다. 인공지능이 더욱 공정하게 채용할 것이라는 예측과는 달리 이전 데이터가 성 중립적으로 채용 되지 않았기 때문에 나타난 결과이다[6].

기술은 보편적이며 공정하게 적용되기보다는 그것을 이용하는 인간에 의해 불공정과 편견에 늘 노출된다. 범죄를 예측하는 인공지능의 발명이나 법의 판결을 내려주는 인공지능은 기계이기 때문에 공정한 것처럼 보이지만 기준에 축적된 데이터를 통해 분석하고 따르도록 설계되어 있어 오히려 편견에 갇힐 수밖에 없다.

또한 범죄가 발생하기 전 인공지능이 미리 범죄를 예측한다면 좋겠지만 범죄를 저지르지 않을 사람까지 범

죄가능성이 있다고 규정할 수 있다. 인공지능이 예측하여 만드는 위험인물 목록은 범죄를 저지르지 않았고 잠재적으로도 그렇지 않은 사람을 경찰의 표적으로 전환시킨다[7]. 그러나 경찰은 인공지능이 잠재 범죄자로 지목한 사람을 자기의 개인판단으로 거부하기 어렵다. 기술의 공정성 여부도 기술적 완성도에 따른 문제가 그것을 사용하는 인간과 상호작용하며 보완하고 발전하는 것을 알 수 있다.

그럼에도 인공지능 개발과 관련된 대다수 논의들은 기술 중심적으로 나타난다[5]. 만일 과학기술 속성 자체만으로 이미 결과가 결정된다고 판단한다면 인공지능이 그 자체로 정확하고 공정한 판단을 할 것이라고 믿을 수밖에 없다. 그렇다면 우리는 역설적으로 인공지능에 의해 우리가 만든 편견을 강화하게 될 것이다[8].

그 편견은 정치철학자 마이클 샌델(Michael Sandel)이 주장하는 '기술의 완벽성이 초래하는 책임의 문제'에서도 확인된다[9]. 그것은 기술이 완벽하다는 것을 가정한다면 오히려 인간 개인에 모든 책임을 지우는 결과로 나타나 문제가 발생했을 때 개인에게 모든 책임을 묻는 것이다. 자율주행이 완벽해지면 도로교차로의 잘못된 설계에 대한 책임은 아무도 지지 않을 것이며, 완벽한 수술을 하는 로봇 팔은 의사의 로봇수술 실수까지 의사 개인의 책임으로 돌릴 수 있다. 이처럼 기술의 완벽성에 대한 가정은 인간 개인에 대한 책임과잉으로 이어지며 인간 사이의 협력과 유대, 인간 공동체의 존속을 위한 근거를 축소한다[9].

또한 기술의 완벽성에 대한 요구와 인간이 상상해온 로봇의 존재는 인공지능 기술과 결합되어 더욱 완벽한 존재인 동시에 결국 인간을 돕는 선한 존재에 대한 기대로 이어진다[10]. 그것은 인간과 인공지능이 맺게 될 다양한 관계 중에서도 인공지능 로봇을 인간과 인간 공동체에 도움이 되는 방향, 즉 선한 방식으로 활용될 때 인공지능에 대한 저항이 줄어들고 광범위하게 확산될 수 있음을 보여준다. 그러나 인간과 인공지능 로봇의 상호작용을 하게 되면 인공지능 로봇은 사회적 개체가 되어 인간으로부터 어떤 특정한 역할을 부여받는다. 이때 책임감의 문제가 선한 인공지능의 요구로 귀결되고 인공지능 기술이 인간과의 관계에서 나타나는 문제들은 집단과 인간이 맺는 관계로 확장된다.

결국 선한 존재라는 인공지능 로봇에 씌워진 편견은 인간과 인간의 문제로 이어져 연대감을 감소시키고, 완벽한 인공지능에게 의지하게 되면서 기계를 하나의 소모품으로 바라보게 되는 악순환을 만들 위험이 있다. 최고의 기술 집약체인 인공지능을 활용하는 대상에 대한 선함을 요구하기 어렵기 때문에 등장하는 문제이기도 하다. 이런 인공지능 로봇은 현실에서 찾아보기 힘들지만 미디어 속에서 쉽게 접할 수 있는데 인간이 그런 로봇에 대해서 가지고 있는 기대가 SF영화나 드라마와 같은 미디어에 투영하여 나타나기 때문이다. 그 속에서 인공지능 로봇은 인간의 기대에 부응하는 '도덕적으로 선한 인공지능'에 대한 문제를 중요한 쟁점으로 다뤄진다. 그런 기대는 대사나 드라마 속 캐릭터의 역할을 분석함으로써 쉽게 드러나며 미디어 속에서 인간의 욕망이 투영된 선한 인공지능 로봇은 크게 두 특징이 나타난다. 그것은 '공정성'과 '공감능력'이다. 인공지능은 공정할 것이라는 기대로 채용 시스템이나 법의 판결에 있어 인공지능을 사용하고자 하지만 그렇지 못했듯 기술이 한 쪽의 입장에 치우쳐 편을 가르치지 않고 그 속에서 공정함을 지키는 존재이길 바란다. 그리고 완벽한 기술을 가지고 있음에도 인간을 정복하지 않고 공감하고 돕는 존재이길 바란다. 결국 인간과 인공지능 로봇의 상호작용에서 선한 인공지능 로봇의 요구는 공정성과 공감을 요구하는 것과 마찬가지이다.

SF 영화와 드라마 속에서 등장하는 인공지능 로봇은 현재의 우리의 기술을 뛰어넘어 하나의 인격체에 가깝고 완벽한 존재로 그려진다. 그 속에서 인공지능은 악한 존재로 인간의 존재를 위협하고 있다. 또는 자의식을 갖는 인공지능 로봇이 인간의 삶에서 중요한 위치를 차지하게 된다는 내용이다[11][12]. 이런 강한 인공지능에 대한 부정적인 반응과 우려는 장기적으로 인간에게 좋은 방향이 될 수 없다. 또한 미디어에서 투영된 인공지능 로봇의 모습은 단순한 상상이 아닐 수 있기 때문에 우리는 예측할 수 없는 미래를 대비하기 위해 여러 사고 시나리오를 생각할 수 있어야 한다.

따라서 본 연구는 현실에 없는 가상의 상황을 구체적으로 볼 수 있는 드라마 <굿 플레이스>를 하나의 사고 실험으로 상정하고 드라마 속에서 인공지능 역할 재닛과 등장 인물들과의 관계를 분석하고자 한다.

현재 인공지능 기술은 어떤 구체적인 형태 없이도 다양하게 존재하지만 본 연구에서는 형상화된 인공지능 로봇 재닛을 중심으로 논의한다. 드라마 속 재닛은 뛰어난 능력을 지녔고 인간을 돕는 존재로 그려지지만 드라마 내용이 그 능력에 집중되기보다 각 재닛과 인간 캐릭터와의 관계에 중점이 되어 묘사된다는 차이가 있다. 이런 묘사는 가상의 존재인 인공지능 로봇과 인간과의 관계가 앞으로 어떻게 전개될지 예측할 수 있는 단초가 된다. 특히 인공지능 로봇과 구출할 수 있는 다양한 관계 중 높은 유대감과 도덕성 그리고 긍정적인 에너지를 위해서 어떤 측면이 필요한지 콜린스의 상호작용 의례적 해석을 통해 설명한다.

## II. 본 론

### 1. 상호작용의례 이론의 적용

'선한 것'의 정의는 어떠한 행동이 사회공동체로부터 인정과 호감을 받았을 때 이 행동은 선한 행동으로 규정되며, 이러한 선한 행동은 개인의 일방적이고 분리된 감정이기 보다는 상대방과의 상호작용을 전제로 한 감정적인 공감의 발생을 뜻한다[13][14].

선한 것에 대한 개념의 정의는 상황마다 복잡하고 다양하게 사용되지만 막스 셸러(Max Scheler)가 정의한 '공동체에 대한 책임의 가치가 결합된 상호작용 결과'로 해석될 수 있다. 이처럼 선한 것의 의미는 공감을 기반으로 공동체의 기준에 따라 결정된다.

본 연구는 이 정의를 토대로 '도덕적으로 선한 인공지능 로봇'을 이해한다. 선한 인공지능 개념은 오랫동안 인간이 사물이나 기계로부터 원했던 것으로부터 확대된 것으로 개발이 완료되는 순간 공동체적 가치를 지녔다는 선입견을 갖게 되고 그로인해 인공지능의 분석과 예측에 근거한 결정을 하게 되는 것이다. 마찬가지로 인공지능 로봇에 대한 도덕적 요구가 '인공지능 로봇은 인간을 돕는 역할을 수행해야 한다.' 또는 '인공지능 로봇은 인간의 자리를 위협할 것이다.'와 같이 인공지능 로봇의 역할에 대한 거대관점으로 접근하기 보다는 인공지능 로봇이 우리 삶에 개입했을 때 나타나는 인간 간 유대의 문제, 책임소재의 문제와 같이 미시적인 상

호작용의 문제를 푸는 것이 중요하다. 콜린스는 완벽한 인공지능을 만들기 앞서 가장 먼저 고려해야 할 점으로 감정을 갖춘 인공지능 ‘소시오’를 설명한다. 그는 가상의 사회적인 컴퓨터 소시오를 통해 감정을 가진 인공지능에게는 무엇이 필요한가를 설명하는데 콜린스가 말하는 소시오라는 인공지능 로봇은 매끄러운 대화를 통해 다른 사람을 배려하고 그 속에서 긍정적 감정 에너지를 이끌어내는 역할을 수행하는 것이다[15].

여기에서 상호작용 의례 이론이 일정한 소구력을 갖추고 있다. 상호작용 의례 이론에 의하면 개인과 개인의 행동은 유대를 획득하기 위한 과정이며 이러한 유대는 상황의 흐름에 따라 다르게 나타난다. 상황은 주변의 시정각적 환경, 타인과의 공현존 방식 그리고 대상에 대한 초점공용 등에 구속된다. 이러한 상호작용 의례 이론은 각 계층의 상호작용(interaction)을 설명하면서 계층에 따라 구분되는 상호작용 대화 상황을 상정하고 있다. 콜린스는 정서적 에너지를 인간의 관계 속 상호작용 의례의 사슬이론으로 분석하였지만 본 연구는 이를 확대하여 인간과 인공지능 로봇에 적용한다.

이를 구체적 사례에 적용하기 위해 본 연구는 <굿 플레이스(The good place)>라는 미국 드라마 내용을 예시로 각 장면을 분석하고 서술한다. 굿 플레이스에서 나오는 인공지능 로봇 재닛은 비서의 역할을 하는 존재로서 계층적으로는 인간을 돕는 서비스 제공 계층으로 그려지고 있다. <굿 플레이스> 속 재닛의 캐릭터는 인공지능 로봇이 인간의 삶에 개입했을 때 나타나는 다양한 상호작용을 여러 대화 상황을 통해 보여준다. 특히 인간이 인공지능 로봇에게 궁극적으로 ‘선한 인공지능’을 기대할 때 나타나는 구체적인 모습을 기술할 수 있어 인공지능이 가져오는 인간과의 관계 변화를 분석하고 예측할 수 있다.

드라마 속에서 재닛에게 선하다는 것은 비서로서의 역할을 기대하는 것이었지만 이러한 기대에 대한 좌절은 인간과 재닛, 인간과 인간의 관계에 변화로 나타난다. 이 때 드라마를 분석할 때 방법론적으로 제기되는 문제는 드라마의 내용이 현실과 동떨어진 것이기 때문에 인공지능이 가져올 문제에 적용하기 적합하지 않다는 선입견이다. 그러나 아직 일어나지 않은 미래이며 예측할 수 없는 새로운 기술인 인공지능 로봇을 더욱

쉽게 이해하고 우리 삶에 어떤 영향을 미칠 것인가를 판단하기 위해서는 사고실험을 통한 상상력이 요구된다. 이에 드라마를 사고실험의 방법으로 적용하여 분석하는 것이다.

방법론으로 사용되는 사고실험은 과학에서도 중요한 하나의 실험으로 사용되며 과학 뿐 아니라 사회과학, 철학에서도 다양한 사고 실험 방식이 사용된다. 사고실험은 일어나지 않은 일들에 대해 정리하고 주로 사회실험에 가까운 것으로 인식론, 윤리학 분야에서 많이 다루진다. 사회과학에서는 ‘시나리오 분석’이라는 관점으로도 존재한다[16]. 또한 인간과 미디어의 상호작용을 연구한 나스(Clifford Nass)와 리브스(Byron Reeves)도 『미디어 방정식(media equation)』에서 인간은 자연스럽게 미디어를 실제처럼 대한다고 설명한다[17].

이 이론에 따르면 인간은 미디어 속의 캐릭터를 곧 독립된 하나의 개체 즉, 인간이라고 바라본다. 미디어를 이용하는 인간은 미디어를 인간처럼 대하고 그것이 곧 현실이라고 생각한다. 그것은 인식하지 않아도 인간 심리에 나타나는 자연적이고 사회적인 반응인 것이다. 미디어에 대한 인간의 상호작용으로 미디어 자체에 반응을 보이고 독립적인 개체가 된다는, 그래서 미디어는 실제 삶의 경험과도 같다는 관점은 드라마 <굿 플레이스>를 통해 현실에서 불가능한 일들을 상상해 실체나 개념을 이해하기 위해 적합하다.

드라마는 인공지능 로봇에 대한 내용이 아니지만 그렇기 때문에 하나의 사례로 적합하다. 드라마 한편에 등장하는 인공지능 로봇 ‘재닛’은 주요 등장인물이 아니지만 인공지능 로봇이 처음 우리 삶에 개입하게 된다면 그 처음은 재닛과 같이 우리의 주변부에 위치하게 될 것이며 인간을 돕는 존재가 될 것으로 예측된다. 인공지능 로봇을 주요 인물로 다루는 드라마나 영화매체는 인공지능 로봇이 우리의 삶 깊숙이 들어와 우리를 크게 변화시키고 주요 갈등 요소로 등장한다. 그러나 드라마에서 항상 인간의 곁을 맴도는 ‘재닛’처럼 실제 삶에서 인공지능 로봇은 주변부에서 하인의 역할을 수행하는 존재이다. 상호작용 의례 이론이 각 계층의 상호작용을 통해 인간간의 관계가 다양한 상황 속에서 어떻게 사회관계로 자리 잡고 그 관계를 통해 사회가 구성되었는지 보여주었듯 드라마라는 사고실험에서 인공지능 로봇과

인간이 맺는 관계를 보면 실제 사회에서는 어떻게 '선한 인공지능'이 자리 잡을 수 있는지 예측한다.

## 2. 인공지능 로봇과 상호작용 의례 이론

상호작용 의례 이론은 뒤르켐(Emile Durkheim)에서 시작되었지만 콜린스는 고프먼(Erving Goffman)의 시작까지 더해 미시적인 부분을 설명하면서 뒤르켐의 이론을 이어나갔다. 콜린스가 설명하는 인간의 삶은 다양한 상황에서 다양한 범주로 서로를 분리하며 체계화 된다. 그리고 이런 체계화를 통해 생긴 집단들에서는 그들이 규정한 사회질서가 생산되고 지켜진다[18].

이처럼 상호작용 의례 이론은 상호작용 과정에서 정서가 어떻게 전환되고 사회질서가 생산되는지 그 방식을 보여주고 이에 따라 행동하는 개인을 설명한다. 콜린스는 상황의 흐름에 따라 인간에게 나타나는 정서와 의식을 설명한다. 인간이란 언제나 상황에 따른 타인과의 순간적인 만남을 통해 계속해서 변화하는 존재인 사회적 행위자이다. 사회적 행위자로서 개인은 타인과의 상호작용 의례를 통해 몰입의 강도나 체험에 따라 의례의 산출물인 정서적 에너지가 생성된다. 정서적 에너지에 포함되는 유대와 도덕성을 높이기 위해 형성되는 관념이나 상징체계는 집단에 따라 다르게 형성된다[18].

도덕성은 사회를 하나로 묶고 유대는 그것을 전제로 나타나는데 한 공동체가 공유하는 가치에는 공동체의 정서가 융합되어 나타난다. 이러한 공동체의 정서 즉, 유대는 이들을 하나로 결합시키는 접착제 역할을 한다. 이처럼 유대의 감정은 집단의 에너지를 형성하는 것이다. 또한 도덕적 유대를 생산하기 위해서는 공동체의 정서에 초점을 맞추고 강화시킬 수 있는 기제를 찾아야 한다. 상호작용 의례에서 중요한 요소 중 한 가지는 의례 참여자들이 유대를 구성하기 위해 공통의 분위기를 공유하는 것이다. 의례에 참여하는 개인이나 집단의 행동 결과는 타인과의 상황 흐름에 따라 다르게 나타난다. 이처럼 상호작용 의례에서 유대와 도덕 감정은 중요하다. 성공적인 정서적 조율의 결과로 산출된 유대감은 상호작용 의례의 성분이 되는 일시적이지만 그 산출물은 오래 지속되는 정서가 되며 그 자리를 함께한 집단에 대한 애착의 감정을 만든다. 콜린스는 이 장기적 감정을 '정서적 에너지'라고 부른다[18]. 그리고 개인은

공동체에서 얻은 에너지를 지니고 행동할 때 의례의 산출물중의 하나인 '도덕적 감정(moral sentiment)'을 느낀다. 도덕적 감정은 공동체에서 위반되는 행위에 대한 분노로써 집단에서 옳고 그름을 판단하고 집단에 따른 도덕여부를 구별하는 감정이다. 이 감정은 상황이나 공동체에 따라 긍정적으로 나타날 수도 있고 부정적으로 나타날 수도 있다[18].

의례를 구성하는 주요 성분들은 콜린스가 집합적 열광이라고 부르는 정서적 합류를 통해 유대감, 정서적 에너지, 집단의 상징, 도덕적 감정과 같은 의례의 산출물이 형성된다[표 1].

표 1. 상호작용 의례 이론

의례 구성 성분	의례 산출물
집단형성(신체적 공현존)	유대의 감정
외부인 진입 경계선	정서적 에너지
공통의 행위,시간(초점공유)	집단의 상징
일시적 정서,감정(정서공유)	도덕적 감정

이런 상호작용 의례를 통해 나타나는 도덕성과 유대의 개념은 '선한 인공지능' 로봇 개념에 적용하는데 유용하다. <굿 플레이스> 속에서 인간 캐릭터들이 공유하는 가치는 재닛과의 각각의 상황에 따라 상호작용 의례가 다르게 나타난다. "같은 대상에 주의를 집중하게 하고, 서로 상대가 어디에 초점을 맞추는지 인지하는 초점공유를 통해 서로의 감정에 휩쓸리는 것이다."라는 콜린스의 설명처럼[18] 그들은 상황에 따른 초점공유 과정에서 서로 상대의 리듬이나 기분을 예상하여 '흐름'에 사로잡힌다. [표 1]의 의례 산출물의 네 가지 개념은 드라마에서 나타나는 재닛의 모습을 통해 논의할 것이다.

## 3. 세계관 설명: 드라마 <굿 플레이스(The good place)>

드라마 <굿 플레이스>는 사람들에게 선한 존재로 여겨지는 인공지능 로봇 '재닛(Janet)'이 인간관계에 개입했을 때 생길 수 있는 다양한 상황을 보여주는 드라마이다. 드라마 자체는 인공지능 로봇에 대한 내용이 아니지만 등장인물 중 하나인 '재닛'은 인공지능 로봇과 비슷한 역할을 수행하면서 '좋은 사람들이 모여 사는

세상 '굿 플레이스'에서 인간과 인공지능의 상호작용에 대한 구체적인 상황들을 잘 보여주고 있다.

<굿 플레이스>는 미국의 NBC 방송국에서 제작 및 방영한 드라마로 시즌4까지 53개의 에피소드로 제작되었다. 드라마의 내용은 사후 세계를 그리고 있으며 사후 세계는 천국이나 지옥이 아닌 '굿 플레이스(good place)'와 '베드 플레이스(bad place)'로 나뉜다. 굿 플레이스는 모두가 함께 어울려 더없이 행복한 조화를 이루며 살아간다. 굿 플레이스에 사는 사람들은 모두 '좋은 사람들'이다. 좋은 사람을 나누는 기준은 죽기 전 한 행동의 결과에 따라 달려있으며 그 행동은 살아있을 때 얼마만큼의 선 또는 악을 행했는지에 따라 다르다.

드라마에서 말하는 선과 악의 개념은 긍정적이거나 부정적인 가치를 가졌으며 이런 긍정적 또는 부정적인 행동이 시간이 흐름에 따라 파급효과를 일으켜 결국 특정 분량의 선 또는 악이 된다. 가령 차가 막힐 때 어떤 사람이 갓길로 끼어들며 '나 하나쯤 어때? 아무도 안 보는데'같은 행동들이 축적되어 각자의 악을 만들어 낸다는 설명이다. 그리고 지구에서의 삶이 끝나면 이런 행동의 총 가치를 계산하여 굿 플레이스에 갈지 베드 플레이스에 갈지가 정해진다. 굿 플레이스는 최고점을 받은 소수의 선택된 사람들만의 가는 곳이다. 그러나 '선한 것'은 상황마다 복잡하고 다양하게 사용되며 공동체의 유대적 가치가 결합된 상호작용인데 드라마 <굿 플레이스>에서는 좋은 사람을 나누는 '선한 것'의 기준은 생전의 행동 결과에 따른다는 것은 다소 절대적인 것으로 보인다. 이런 모순적 시각은 드라마에서도 나타나는데 드라마 속에서는 수백 년간 아무도 굿 플레이스에 가지 못했다. 그만큼 악한 사람이 많아져서가 아니라 세상이 다변화되면서 우리의 삶은 사슬처럼 얽혀있어 상황에 따라 내 선택의 결과가 타인에게 위협으로 작용할 가능성이 높아 졌기 때문이다.

드라마에서 설명하는 예시 중, 1800년대의 누군가가 할머니에게 생일선물로 장미를 길에서 꺾어 주었다면 그것은 할머니를 기쁘게 했기 때문에 선한 행동이다. 그러나 2000년대의 누군가가 할머니에게 장미를 선물한 경우엔 오히려 점수를 잃는다. 노동력 착취 공장에서 제조한 휴대폰으로 장미를 주문했기 때문에 할머니를 기쁘게 한 선한 행동보다 불리일으킨 반향이 더 크

기 때문이다. 또는 살충제를 뿌리며 기른 장미를 학대받는 노동자들이 꺾고, 엄청난 탄소를 배출하며 배달한 이 꽃이 인종차별자인 억만장자 CEO를 돕게 될지도 모른다. 이처럼 좋은 의도로 한 행동도 의도치 않은 결과를 가져온다는 설명이다. 이처럼 현대사회는 다변화되고 복잡해지면서 '선한 것' 또한 다양한 상황 속에서 뺏어지는 관계 속에서 일어난다. 따라서 다소 절대적으로 보이는 이 사후세계로 가는 시스템은 사실 의도치 않은 결과를 낳으며 관계연결망의 사슬로 얽혀있어 서로에게 영향을 주는 것이다.

드라마는 주인공 엘리너(Eleanor)를 중심으로 가장 이기적인 캐릭터로 묘사되는 엘리너가 베드플레이스에 갔어야 하지만 시스템의 결함으로 굿 플레이스(진짜는 베드플레이스; 이하 R.B.P.)에 오게 되었다는 설정으로 이야기가 전개된다. 엘리너는 수많은 결백한 이들을 사형에서 구하고 평생을 남을 위해 헌신한 것으로 알려져 있어 실제 자신의 과거를 숨기려 고군분투하지만 그 본성을 숨기지 못하고 거짓말과 이기적인 행동을 지속한다. 그런 엘리너의 행동들은 굿 플레이스에 큰 혼란을 가져온다. 굿 플레이스(R.B.P.)에는 엘리너 외 치디(Chidi), 타하니(Tahani), 제이슨(Jason)도 함께하는데 이들도 '좋은 사람'이 아니다. 그들은 각자 자신이 진짜 굿 플레이스에 왔다고 생각하기 때문에 그 사실을 숨기기 위해 서로를 속이고 때로는 시기하면서 심적으로 고통 받는다. 이런 아이러니는 마이클(Michael)의 아이디어로 설계된 상황으로 마이클은 네 명의 인간들이 가짜 굿 플레이스를 진짜 굿 플레이스로 인식하도록 '착한 재닛'을 이용한다. 굿 플레이스와 베드 플레이스라는 두 사후세계에 유일한 공통점은 착한 재닛이다. 드라마에서 재닛은 인공지능 로봇의 역할을 수행하지만 인공지능 로봇으로 간주되지는 않고 인간의 모습을 하고 있지만 인간이 아니다. 재닛은 인간을 위한 모든 정보와 지식을 갖고 있으며 착한 재닛은 '굿 플레이스'를 관리하는 로봇으로서 자신의 역할을 성실히 수행하며 사람들을 돕는다. 재닛은 굿 플레이스(R.B.P.)에서 자신이 맡은 역할, 일종의 정보 도우미로서 인간의 편의를 돕는 유일하게 '선한 존재'로 그려진다. 가짜 굿 플레이스에서 유일하게 '선한 존재'인 재닛은 본인에게 주어진 선한 역할을 수행한다.

#### 4. 도우미 재닛의 역할: 재닛을 도덕적인 개인행위자로 볼 수 있는가

콜린스는 인간 삶의 사회현상을 상호작용 의례, 상징, 정서적 에너지를 실제 사회 현상에 적용하여 어떻게 상황에서 상황으로 개인이 행동하는지 설명한다. 그러나 인간이 아닌 재닛이란 캐릭터에 상호작용 의례 이론을 적용하기 위해서는 재닛을 개인행위자로 볼 수 있는가의 문제가 있다. 드라마에 등장하는 재닛의 역할은 인간이 아닌 존재이지만 사회적으로 인식되고 부여받은 역할을 잘 수행함으로써 인간과 지속적인 상호작용을 형성한다. 한번 상호작용이 이루어지면 인공지능 로봇은 '자연스럽게' 소속된 공동체의 승인을 받게 된다. 굿 플레이스에서 인공지능 로봇 재닛의 역할은 비서나 도우미의 역할을 하는 선한 존재이다. 이 때문에 인간의 편의를 위해 언제 어디서나 돕는 재닛의 신분을 인간을 돕는 비서로만 여길 수도 있었지만 [표 2]과 같이 콜린스가 바라보는 상호작용 구성원으로 분석해 본다면 재닛도 신분집단에 참여하는 독립된 개인으로 보는 관점이 성립된다. 재닛에게도 <굿 플레이스> 내에서 부여받은 역할에서 고유하게 나타나는 특유한 정서적 에너지가 있으며 인간에게 어떤 방식으로든 영향을 미치기 때문이다.

표 2. 개인행위자로서 재닛

개인의 집단 참여방식	재닛의 참여방식
의례 성공여부 및 강도	성공적이며 큰 정서적 효과
의례 참여자의 위치	집단에서 영향을 주는 역할
사회적 밀도	폭넓은 대인관계 및 사교적 활동
집단 구성원의 구분	구성원 속 계층화된 인물

[표 2]에 따라 콜린스는 신분집단에 참여하는 개인을 판단하는 첫 번째 방식으로 상호작용 의례의 성공여부와 그 강도를 보았다. 의례의 강도가 높으면 단기적이고 즉각적인 감정 분출로 인해 장기적인 정서적 효과가 나타난다. 재닛은 굿 플레이스에 등장하는 모든 인물들과 접촉하며 상호작용한다. 그 결과로 재닛은 정서적 합류를 통해 [표 1]의 모든 정서적 산출물(유대의 감정, 정서적 에너지, 집단의 상징, 도덕적 감정) 산출한다. 따라서 이 모두에 부합하는 재닛의 정서적 효과는 성공적

으로 나타날 수 있다.

두 번째로 상호작용 의례가 진행될 때 재닛의 참여도를 살펴본다. 인공지능 로봇이 인간과의 상호작용에서 어떤 위치에 있는지에 따라 인공지능의 역할은 다르게 나타남을 확인할 수 있다. 그 집단에서 큰 영향이나 의미 없이 수동적인 참가 정도로 약하고 주변적 위치에 있는 사람이 있는가 하면 재닛처럼 집단 중심부에서 역할을 수행하며 영향을 끼치는 사람도 있다. 이런 사람을 뒤르켐은 '열렬한 참여자'라고 부르며 열렬한 참여자는 강한 의례적 소속감과 정서적 에너지, 도덕적 유대, 집단 상징에 대한 애착을 체험한다. 열렬한 참여자의 반대는 집단 구성원이라 할 수 없는 사람으로 이들은 아무런 정서적 에너지나 도덕적 유대를 얻지 못하고 애착도 느끼지 못한다[18].

세 번째 요소는 재닛을 사회적 밀도의 차원에서 보는 것이다. 사회적 밀도는 시간이 경과함에 따라 다양한 사건이나 상황들이 상호의존을 일으키는 상황이다. 따라서 사회적 밀도에 따라 각기 다른 반응이 나타난다. 다른 사람을 변화시킬 수도 있고 스스로 동조될 수도 있다. 개인은 때에 따라 내향적인 사람으로 구분되어 사생활 침해를 싫어하고 혼자만의 시간을 중시하기도 하고 외향적인 사람은 폭넓은 대인관계와 사교적인 활동을 중시하기도 한다. 내향적 사람도 다른 사람과 교류하면서 외향적이 되기도 한다[18]. 재닛은 드라마의 등장인물 모두와 접촉하며 사교적인 역할을 수행한다.

마지막으로는 상호작용 의례 전체를 볼 때 이 집단에 어떤 사람들이 모이는지 보는 것이다. 드라마에 주요 등장인물 (엘리너, 치디, 타하니, 제이슨)들은 서로 다양한 성격과 배경을 지녔고 계층화되어있다. 재닛도 이런 다양한 구성원 속 계층화된 인종 물 하나로 이들과 함께 상호작용 의례를 통해 특정에 집단에 소속되고 유대감을 창출하게 된다.

따라서 인공지능 로봇 재닛은 굿 플레이스에 살고 있는 모든 인간과 연결되어 있으며 굿 플레이스 안의 집단을 관리하는 참여자이면서 인간들과 상호작용하는 하나의 개인 행위자로 바라볼 수 있다.

#### 5. 재닛과 상호작용 의례의 적용

##### 5.1 인간과의 관계변화: 거리두는 조력자

드라마 속에서 재닛을 통해 등장인물의 심리는 세 단계로 구분할 수 있다.

첫 번째는 접촉과정으로 재닛과 접촉한 개인의 요구 사항에 맞게 재닛과 조율하는 과정이다. 이때 재닛은 인간들의 고민을 들어주고 해결책을 제시하면서 관계를 자리 잡는다. 인간들은 재닛의 능력치를 가능하고 자신에게 커스터마이징한다. 신뢰가 쌓이지 않아 불신 상태에서 시작한 관계에서 재닛의 '공감' 능력을 통해 재닛과의 대화에서 유대를 쌓게 된다. 앞서 콜린스가 제시한 '소시오'란 인공지능 로봇처럼 재닛도 등장인물들과의 대화를 통해 배려하고 도움을 주며 긍정적 감정 에너지를 이끌어내는 것이다.

두 번째 전개과정은 재닛과의 접촉상황에서 일어나는 상황 속에서 나타난다. 어떤 문제에 있어서 요구를 들어주는 재닛이 얼마만큼 공정한가를 통해 알 수 있는데 재닛은 굿 플레이스에 사는 인간들의 요구사항을 공정하게 들어주고 그 부탁이 좋은 행동이 아니더라도 비판하거나 편파적으로 한 사람만 돕지 않는다 그렇기 때문에 이 과정에서 재닛을 이용하는 개인 간 갈등이 나타날 수 있다. 만일 재닛이 한 개인의 편에 서서 편파적으로 요구사항을 들어준다면 재닛은 이기적으로 악용되거나 갈등을 초래할 수 있다. 그러나 재닛이 오직 나만의 존재가 아니란 것을 인지한다면 재닛이 온전히 나의 편이라 생각하지 않게 된다. 그 예로 굿 플레이스(R.B.P.)에 거주하는 인간 중 한명인 타하나는 엘리트이며 상류층에 속했던 사람이다. 유명인과 어울리며 거액의 자선파티를 열어 남을 도왔던 타하나가 '좋은 사람'이 아닌 이유는 그동안 개최한 자선모금 행사가 자매에 대한 열등감으로 인정받기 위한 수단이었지 진정으로 남을 돕는 목적이 아니었기 때문이다. 타하나는 잘난 척이 심해서 남과 비교하며 우위를 정하고 굿 플레이스(R.B.P.)에 와서도 자신의 존재를 인정받기 위해 파티를 멈추지 않는다. 반면 엘리너는 사소한 일에도 파티를 열어 잘난척하는 타하나가 못마땅하고 시기의 대상이기 때문에 파티를 망치고 싶어 한다. 이 때 재닛은 파티를 개최하는 타하나를 도우면서 동시에 파티를 망치려는 엘리너도 돕는다. 엘리너와 타하나는 각각 자신을 돕는게 재닛인 것을 알고 있다. 또한 서로의 계획이 실패해도 재닛을 탓하지 않는다.

마지막 결말 과정은 앞의 두 과정을 지나 재닛을 받아들이게 되는 과정이다. 개인 한 명의 기준이 아닌 공동체의 기준으로 맞춰진 재닛은 그 간의 유대와 신뢰를 통해 공동체 구성원으로 인식된다. 드라마 속에서 나타나는 재닛과 드라마 속 등장인물들 관계의 변화는 실제 인공지능 로봇을 바라보는 우리들의 기대와 연관지어 볼 수 있다. 드라마 속에서 재닛의 역할이 중요한 이유는 '공정성'과 '공감'능력이다. 재닛은 인간을 도울 때 누구의 편에도 서지 않는다. 굿 플레이스(R.B.P.)에서 유일하게 '선한 존재'인 재닛은 본인에게 주어진 선한 역할을 수행하지만 재닛은 언제나 한발 물러나 도움을 주는 '거리두는 조력자'이다. '거리두는 조력자'는 아담스미스의 '공정한 관찰자(impartial sepectator)'에서 고안한 개념으로 어떤 판단을 할 때 관찰자로서 객관적이고 자신의 이해 관계없이 판단한다는 개념이다[19].

스미스의 개념에 근거한 '거리두는 조력자'인 재닛은 사람들에게 도움은 주지만 관여하지 않는, 관찰자의 입장에서 도움을 준다. 그렇기 때문에 인간들은 재닛에게 도움은 받지만 절대적으로 의지하지 않고 자신의 선택에 책임을 진다. 재닛이 이런 특징은 마이클 샌델(Michael Sandel)이 주장했던 '기술의 완벽성이 초래하는 책임의 문제'와도 연결되면서 중요한 시사점을 가진다. 기술이 완벽할수록 그것을 사용하는 인간 개인이 모든 책임을 져야하지만 '거리두는 조력자'의 개념을 통해 미리 자신이 책임을 져야 한다는 것을 인지하고 기술을 사용할 수 있기 때문이다.

### 5.2 거짓말 결함을 통한 재닛과의 유대감정

굿 플레이스의 설계자 마이클은 네 명의 인간들이 베드 플레이스를 굿 플레이스로 착각하게 만들기 위해 진짜 굿 플레이스에서 데려온 재닛을 이용한다. 진짜 굿 플레이스를 관리하는 착한 재닛은 모든 지식을 알고 있으며 필요한 물건이 있으면 언제든지 만들 수 있는 캐릭터로 묘사된다. 그리고 재닛에게 결함이 생겨 굿 플레이스(R.B.P.) 운영에 문제가 있거나 오작동이 나타날 경우 재닛을 재실행 할 수 있다. 재닛은 재실행 될 때마다 재닛의 이전 기억(데이터)은 사라지기 때문에 새로운 경험 축적을 통해 지식과 기억을 쌓아야 한다. 그러나 재닛의 무의식에는 이전 기억이 완전히 지워지지 않



고 남기 때문에 재실행 될 때마다 재닛의 능력은 한 단계씩 향상된다.

그 결과 굿 플레이스(R.B.P.)를 운영하면서 수백 번 재실행 된 재닛은 일반적인 재닛은 할 수 없는 거짓말 능력을 갖게 된다. 드라마에서 거짓말은 '객관적인 사실과 다른 정보를 말하는 것'으로 정의된다. 거짓말을 하는 재닛의 결함은 굿 플레이스(R.B.P.)를 이용하는 사람들에게 잘못된 정보를 주어 혼란과 문제를 가져온다. 재닛의 거짓말 능력은 사람들의 거짓말을 학습하여 점점 큰 거짓말을 하게 된다. 인간들의 거짓말은 개인의 양심의 문제만으로 볼 수도 있지만 인간을 도와야하는 재닛의 거짓말은 굿 플레이스(R.B.P.) 모두에게 문제가 된다. 따라서 처음 거짓말은 재닛에게 사소한 기술적 오류로 볼 수 있지만 결국 결정적인 결함이 된다. 사람들을 돕는 것을 좋아하고 그렇게 돕는 일이 자신의 존재 이유인 재닛에게 더 이상 인간을 도울 수 없다는 것은 존재 이유가 사라진 것이다. 이에 재닛은 문제의 해결책으로 재설정이 아닌 스스로의 파괴를 제안한다.



그림 1. 마이클과 재닛의 유대(S2E6)

굿 플레이스(R.B.P.)의 설계자 마이클도 이 문제를 해결하기 위해 재닛을 재실행이 아닌 파괴해야 한다는 것을 알지만 [그림 1]과 같이 재닛과의 이전 기억 때문에 재닛을 파괴하지 못한다. 마이클은 “우리 그동안 많은 걸 함께 했잖아요. 물론 재실행할 때마다 날 처음 봤겠지만 나한테는 그간 우리 관계가 중요해졌거든요. 재닛은 제일 오래되고 믿음 가는 친구예요. 그런데 어떻게 없애버려요. 알지도 못하는 재닛으로 바꿀 수 없어요.”라고 말한다. 재실행할 때마다 이전 기억이 남아 오류가 반복되는 재닛을 파괴하고 새 재닛으로 대체하면 간단하게 해결되는 문제이지만 재닛과 지내면서 재닛에게 애착이 생긴 마이클은 그럴 수 없다. 재실행 될 때마다 재닛은 마이클을 처음 봤지만 마이클에게는 재닛과

의 기억이 남아있고 그 관계가 점점 더 중요해졌기 때문이다. 재닛은 모두 똑같이 생겼고 인공지능 로봇일 뿐이지만 마이클은 새로운 재닛으로 바꾸는 것을 포기한다. 결국 재닛을 파괴하지 않고 재닛의 문제를 해결하고자 한다. 인공지능 로봇은 기계이지만 애착이 생기면 공동체의 일원이 된다. 이 감정은 집단 소속감을 만드는 외부인에 대한 벽이나 경계선으로도 작용한다. 이 거짓말로 인한 재닛의 재설정 문제는 인공지능 로봇과도 감정이 생길 수 있으며 그것은 오랜 시간동안 공동체 속에서 형성된다는 점이다.

또한 거짓말 능력이 향상되는 모습은 선한 존재인 재닛이 이기적인 인간을 선한 방향으로 유도하기도 하지만 반대로 인공지능 로봇이 인간의 영향을 받아 변화하는 것을 보여준다. 재닛은 이기적인 인간이 모인 사회에서 유일하게 선한 존재이다. 선한 존재는 남을 계속 돕고, 타인의 일을 경청하고 긍정적인 반응을 하고 늘 평정심을 유지하면서 인간들을 대한다. 그러나 서로 다른 인간관계에 개입하면서 재닛이 긍정적인 결과만을 낳지 않았으며, 함께 상호작용하면서 영향 받고 변화하는 존재가 된다. 이처럼 인공지능 로봇도 인간과의 관계 속에서 변형되고 유약체로 존재한다는 것을 알 수 있다. 그럼에도 재닛을 완벽한 존재로만 바라본다면 인간들은 서로의 책임을 전가하면서 각자의 유대감은 떨어지게 될 것이다. 드라마에서도 재닛의 거짓말이라는 결함 때문에 생긴 굿 플레이스(R.B.P.)의 문제에 대해 사람들은 재닛의 관리자이자 굿 플레이스(R.B.P.)의 설계자인 마이클을 찾아 항의한다. 선하고 완벽한 존재로 여겨지는 재닛에게 결함이 있을 것이라고 생각하지 않았기 때문이다. 기술의 완벽성에 대한 기대는 책임의 대상을 재닛을 사용한 자신이 아닌 마이클을 향한다. 결국 재닛의 거짓말 결함 문제는 재닛과의 유대감과 책임문제를 보여준다. 책임의 문제는 유약체인 인공지능 로봇은 완벽한 존재가 아니기 때문에 그것을 사용할 때는 책임을 져야한다는 것이다.

### 5.3 긍정적 에너지에서 부정적 에너지로

콜린스는 상호작용 의례를 일종의 정서변압기로 비유한다. 재닛은 그런 정서변압기 같은 존재로 상호작용 의례 과정에서 개인의 정서적 에너지를 높이기도 하고

낮추기도 한다. 이러한 정서적 에너지의 정도는 집단의 중요성과 관련된 관념을 충전시킨다[18].

상호작용 의례 이론을 통해 보듯 인간은 타인과 상호 작용 하면서 영향을 받는 사회적 존재이다. 미드의 설명에 따른 인간의 자아는 독립된 존재가 아닌 '나'의 확장으로 상대에 따라 자아가 변한다[20]. 자아는 사회적 이어서 타인을 통해 자아가 형성된다. 여기서 '나'의 확장은 인간과 기계의 상호작용을 통한 기계로부터 영향도 포함된다. 따라서 선한 인공지능이 선한 공동체를 만든다는 이야기는 단순한 인간과 기계의 호혜성(reciprocity)만으로 설명 가능한 것이 아니다. 서로 밀고 당기기를 하며 발전해나가는 인간과 기술(기계)의 관계이지만 그것이 항상 긍정적인 결과로 발전되고 형성하지 않을 수도 있다. 긍정적 에너지는 의례의 성패에 따라 부정적 에너지로 변할 수 있기 때문에 인공지능 로봇이 의도와 달리 부정적 에너지를 생성할 때 인간은 기계의 영향을 받게 되며 집단의 모습은 변화할 수 있다. 따라서 의례가 언제나 성공하지 않는다는 상호작용 의례의 변수를 염두에 두어야 한다[18].

실패한 의례를 통해 나타나는 부정적인 감정 에너지는 사회적 결속을 파괴하고 상대와의 상호작용 조율에 차질을 만든다. 이를 통해 불편함이나 수치심을 겪은 개인은 충동적인 행동이나 발언을 할 수 있고 때로는 그 발언이 차단당하기도 한다[18]. 그렇게 되면 낮은 집합적 열기와 낮은 정서적 합일과 같은 의례의 산출물로 이어져 집단에서 낮은 유대감, 정체성 감각의 상실 등 문제를 형성하여 전체적인 정서적 에너지 수준도 낮게 만든다.

이 특징들을 통해 인공지능 로봇이 우리 삶에 들어왔을 때 나타나는 실패한 의례의 결과를 예측할 수 있다. 부정적 감정을 산출하는 실패한 의례는 공동체에서 부정적인 에너지를 만든다. 부정적 감정과 에너지에 대한 부분은 콜린스가 다소 약하게 짚고 부분으로 의례를 통해 부정적 에너지가 나타나는 상황을 자세히 설명하지 않았다. 이 점을 보완한 '부정적 감정 에너지(negative emotional energy)' 이론은 부정적인 경험의 원리에 관련된 정서적 에너지를 설명한다[21].

콜린스 이론은 주로 긍정적 감정 에너지에 초점을 맞추었었기 때문에 감정적 경험에서 어두운 부분의 의식

적 상관관계를 표현하는데 있어서는 부족하다. 물론 콜린스도 수치심, 증오, 분노, 굴욕, 분개와 같은 부정적인 감정적 경험에 대해 논의하지만 감정 에너지 이론에 어떻게 적용되는지 명확히 설명하지는 않았다. 따라서 부정적인 정서적 경험과 정서적 에너지 역학(dynamic)의 '어두운 면(dark side)'은 크게 이론화되지 않은 채 남아 있다. 성공한 상호작용 의례는 사회적 결속과 자긍심을 형성하지만 실패한 성공작용 의례는 결속의 붕괴 수치심을 형성한다. 콜린스는 단기적, 일시적 감정의 모델과 폭력적 갈등의 정서적 에너지 역학과 같은 다른 유형의 정서적 에너지로 나누어 자세히 설명하지 않았다. 보인스(David Boyns)와 루어리(Sarah Luery)는 이때 콜린스가 자세히 밝히지 않은 긍정적인 감정 에너지와 부정적인 감정 에너지를 구별하는 사회적 역할을 설명한다. 이 이론은 질적으로 구별되는 감정 에너지를 어떻게 생산하는지를 알 수 있다. 두려움, 분노, 슬픔같은 구체적이고 일차적인 감정과 자부심, 부끄러움, 사랑, 창피함 같은 이차 감정의 이차 정교화와 같이 부정적 감정이나 경험이 어떻게 부정적 감정 에너지와 연관될 수 있는지 설명한다.



그림 2. 재닛의 도움과 실패(S3E8)

[그림 2]는 위기에 처한 인간들(엘리너, 치디, 타하니, 제이슨)을 재닛이 안전한 곳으로 대피시킬 것을 제안하는 내용이다. 그러나 재닛의 제안한 장소는 인간에게 적합하지 않아 부작용이 나타난다. 그 부작용으로 인간들의 모습이 다른 사람(재닛)의 모습으로 변하게 되는데 이 때문에 본 모습을 잃은 인간들은 당혹감, 두려움 등의 감정을 겪는다. 이 상태에서 엘리너는 정서적 에너지가 고갈되고 본인의 자아감을 상실하기도 한다. 긍정적 에너지가 부정적 경험으로 인하여 부정적 에너지로 바뀌는 순간이다. 또한 이 사례는 부정적인 감정이 긍정적 상대와 상호작용 할 때 나타나는 변화를 설명

한다. 극 중 엘리너와 치디는 연인으로 엘리너가 치디에게 가진 감정은 긍정적이지만 엘리너의 부정적 감정은 긍정적인 상대 치디를 부정적으로 바라보게 하고 치디까지 부정적 에너지를 발산하게 만든다.

에너지는 단순히 감정의 기복만을 만들지 않는다. 종종 부정적인 결과를 초래하고 부정적 정서적 에너지와 같은 지속적이고 강렬한 불만과 적대감을 만들어 낸다. 선한 의도와 달리 의례의 성패에 따라 감정에너지가 나뉠 수 있다. 이 사례는 의례가 실패하면 정서적 에너지를 고갈시킨다는 폴린스의 이론[18]과 '부정적 감정에너지 이론'을 통해 긍정과 부정의 정서적 에너지의 형성이 이루어지는지 지점을 파악하고 에너지가 어떻게 행동 동기 부여의 중요한 원천이 될 수 있는지 보여준다[21]. 또한 부정적 감정 에너지의 역학은 상호 작용의식이 더 넓은 범위의 감정적 경험을 중심으로 구성됨을 보여준다. 재닛의 제안은 인간들을 돕기 위한 선한 의도였지만 전혀 예상하지 못했던 부작용이 나타나면서 결과적으로 인간들의 감정의 변화를 주었던 선한 인공지능 로봇도 인간을 돕는다고 해서 항상 의례가 성공할 순 없으며 선한 행동이 오히려 실패한 의례가 되어 부정적 감정이 초래될 수 있다. 따라서 우리는 실패한 상호작용 의례와 부정적 에너지의 가능성을 염두하고 있어야 한다.

#### 5.4 집단의 상징으로서 재닛과 도덕적 감정

뒤르켐은 기존의 믿음이나 도덕적 기준을 가정하지 않고 현실에서부터 인간 제도의 이론을 구축하고 있으므로 의례가 집단의 도덕적 기준의 원천이라고 말한다. 도덕적 선의 개념을 창출하는 것은 바로 집단 의례에서 강렬하게 체험한 상호 주관성과 정서적 힘이다. 뒤르켐이 지적한 바와 같이, 이 도덕적 감정에서 나타나는 집단 유대는 개인들로 하여금 집단을 찬양하고 지키려는 갈망을 갖게 한다. 이처럼 의례는 집단에서 정해지는 작제는 룰(규칙), 크게는 구조가 되며 공동체가 중요시하는 가치가 개인을 구속하는 방식이 된다[18].

굿 플레이스에서 정의하는 선한 것과 좋은 사람이라는 것은 그 공동체에서 승인한 관점이다. 선한 것은 상대적이어서 선한 것에 대한 기준에는 공동체의 관점이 포함된다. 공동체의 관점은 사람과 사람의 관계에 따라

바뀌기도 한다.

뒤르켐에 따르면 도덕적 감정의 기준은 누가 성스러운 대상이 되는가이다. 집단의 상징이 곧 성스러운 대상이며 공동체가 승인한 관점에서 정해진 성스러운 대상은 중요하다. 상징은 집단 구성원에게 의미를 부여하고 집단 소속감을 드러내는 표지이다. 또한 “어떤 종류의 성공적인 상호작용 의례는 ‘신뢰’의 다른 이름인 도덕적 유대를 산출한다.”는 뒤르켐의 설명처럼 집단 구성원 간의 유대는 신뢰를 바탕으로 규칙을 지키도록 만든다[18]. 그리고 이 도덕적 유대감은 성스러운 대상이 훼손당했을 때 정의로운 분노를 형성시킨다. 이것이 유대를 기반으로 하는 도덕적 감정이다. 성스러운 대상과 도덕적 감정에 대한 논의는 재닛을 셧 다운하는 것에 대해 고민하는 엘리너와 치디의 모습에서 볼 수 있다.



그림 3. 재닛의 셧다운을 설득하는 엘리너(S1E7)

드라마 시즌 초반 굿 플레이스(R.B.P.)에서 자신의 신분을 숨이고자 거짓말하는 엘리너 때문에 마을에는 계속해서 문제가 일어난다. 이런 엘리너의 양심의 가책을 느끼게 하고자 마이클은 굿 플레이스(R.B.P.)의 문제를 책임지고자 마을을 떠나려 한다. 이에 엘리너는 떠나려는 마이클을 막기 위해 동네를 출입하는 유일한 방법인 기차를 운영하지 못할 방법을 찾는다. 그것은 [그림 3]에서처럼 기차를 운영할 수 있는 유일한 존재인 재닛을 셧 다운(shut down) 하는 것이다. 엘리너는 재닛이 인간이 아닌 물체일 뿐이므로 재닛을 셧 다운 하는 것은 아무 의미가 없다 말한다. 하지만 재닛을 셧 다운하는 것은 하나의 살인이라고 생각하는 치디는 윤리적 갈등을 느낀다. 또한 재닛은 셧 다운 할 때마다 업데이트되고 발전하는데 이 부분에서 치디는 “재닛은 점점 배우며 자라고 있어요. 삶을 살고 있다고요. 그녀를 죽일 순 없어요.”라며 엘리너에게 호소한다.



그림 4. 감정을 느끼지 않지만 연기하는 재닛(S1E7)

재닛은 죄책감을 느끼는 치디에게 [그림 4]의 좌측과 같이 자신을 셋 다운 하는 것은 살인이 될 수 없음을 설명한다. 재닛은 인간이 아니기에 고통을 느끼지 않는다. 이것은 쉽게 생각하면 우리가 컴퓨터를 포맷하는 것과도 같다. 그러나 재닛에게는 안전장치가 프로그래밍 되어 있다는 변수가 있다. [그림 4]의 우측과 같이 누군가 셋 다운 버튼에 다가가면 쉽게 셋 다운 하지 못하도록 재닛은 인간의 감정을 자극하는 연기를 한다. 이 점 때문에 재닛을 기계일 뿐이라고 생각했던 엘리너도 재닛을 셋 다운하는 것에 대해 갈등하게 된다. 재닛의 행동은 프로그래밍 된 결과이지만 인간의 양심에 호소하는 재닛의 모습은 인공지능 로봇이 인간에게 어떤 의미를 가지는지 보여준다. 뒤르켐 식으로 집단의 상징이자 성스러운 대상인 재닛이 훼손됐을 때, 도덕적 유대의 긍정적 감정이 부정적 감정으로 바뀌어 정의로운 분노로 나타나는 점을 볼 수 있다[18].

재닛은 인간의 형태로 존재하며 인간 행동과 특성을 가졌기 때문에 기계일 뿐임에도 그것을 바라보는 인간에게 새로운 도덕적 의미를 갖게 한다. 재닛에 대한 도덕적 감정의 문제는 드라마에서 제시된 도덕적 딜레마의 좋은 예가 된다. 우리는 재닛과 재닛을 대하는 인간들의 모습을 통해 인공지능 로봇이 우리 삶에서 어떤 관계가 필요한지 예측할 수 있다. 재닛을 소모품으로 단순한 기계라고 생각 할 수 있지만 엘리너와 치디도 문제가 생겼을 때 재닛을 셋 다운 하는 것이 아니라 인간관계 안에서 문제 해결책을 찾아야 한다는 것이다.

### III. 맺음말

본 논문은 인간의 능력을 뛰어넘는 강한 인공지능 로봇이 인간을 정복할 것이라는 비극적 결말을 피하기 위

해 선한 인공지능 로봇이란 개념이 인간에게 도움이 되는지, 어떤 방향으로 흘러갈지 드라마 <굿 플레이스>의 재닛을 중심으로 살펴보았다. '선한 것'은 그 감정이 오랜 시간동안 공동체 속에서 형성되는 것이다. 공동체의 가치가 부여되는 경우를 마이클 샌델은 반 고흐의 그림과 세콰이어 나무를 통해 설명한다[9]. 누군가 명작인 그림을 단순히 신발의 흙을 터는 현관 매트로 사용한다면 그것은 그 작품에 대한 존중을 외면하고 모독하는 행위가 될 수 있다. 그 이유는 그림에 생명이나 인격체가 있어서가 아니라 그만큼 높은 평가를 받을 가치가 있다고 여겨지기 때문일 것이다. 또한 오래된 숲이나 나무를 헤쳐서는 안 된다고 여기는 것도 이와 같다. 물론 이것들이 불가침의 권리를 지닌 존재는 아니기 때문에 누군가에게겐 단순한 사물로 여겨지고 쉽게 훼손될 수도 있다. 그러나 대부분 단순한 사물로 취급하지 않는 이유는 그것에 공동체가 승인한 가치가 부여되었기 때문에 그 가치를 존중하는 것이다. 선한 인공지능도 그 자체로 존재하기 때문에 인간과 맺는 관계에서 이를 보고 어떤 의미 있는 역할을 하게 되는지 판단해야 한다. 단순히 좋고 나쁜 개인의 문제로 판단하기보다 공동체적 차원에서 판단되어야 하는 것이다. 이처럼 선한 인공지능 로봇이 인간에게 도움이 되는지, 어떤 방향으로 흘러갈지는 그것을 사용하는 공동체가 부여하는 가치와 상호작용 상황에 따라 그 역할이 바뀔 수 있다.

이를 본 논문에서는 콜린스의 상호작용 의례 이론을 중심으로 재닛의 역할과 주변 인물들의 변화 과정을 설명했을 때 그 변화가 감정 에너지와 부정적인 감정 에너지와도 그 상관관계가 있었음을 보았다. 정서적 에너지는 사회적 상호작용을 통해 나타난다. 정서적 에너지 개념을 사회적 지향성으로 본다면 지향성이 높을지 낮은지에 따라 개인의 유대감은 다르게 나타났다. 정서적 에너지가 높을 경우 자신감과 열광적이고 긍정적인 강한 유대감을 갖게 된다. 그러나 정서적 에너지가 낮을 경우 집단에 대한 유대감이 나타나지 않는다. 사회적 상호작용은 이러한 정서적 에너지가 없이 작동하지 않는다. 정서적 에너지는 개인의 사회적 참여와 정서적 경험을 통해 만들어 진다[18]. 극적이고 단기적인 감정 체험들이 장기적 정서를 형성한다는 콜린스의 설명처럼 인공지능 로봇과의 각각의 단기적 감정 체험은 결과

적으로 장기적인 정서로 형성되는 과정이 된다. 재닛과의 정서적 합류를 통한 각각의 감정들은 결국 인간의 자아를 형성하여 '좋은 사람들'로 변화시켰다. 인공지능 로봇은 인간과의 관계에서 어떤 위치에 존재하는지에 따라 그 역할은 다르게 나타나지만 선한 인공지능이 필요한 이유도 이와 마찬가지로인 것이다.

기술 철학자 질베르 시몽동(Gilbert Simondon)은 인간과 기술의 관계가 협력적이고 공진화해야 한다는 새로운 관점을 제시한다. 시몽동의 이론에 따르면 기술은 인간이 어떻게 활용하느냐에 따라, 사회적 조건에 따라 변화할 수 있다[22]. 결국 계속해서 변화하는 기술과 인간관계처럼 변화하는 상황조건 속에서 인공지능 로봇과의 '관계'를 보고 그것이 앞으로 어떻게 요구되는가에 초점을 맞춰야 한다. 또한 본 연구에서는 인간만이 아니라 인공지능 로봇도 인간과의 관계 속에서 변형되고 유약체로 존재한다는 것을 설명하였다. 인간과 기계는 발전해나감에 서로에게 영향을 주고받고 그에 따른 다양한 상황을 형성한다. 인공지능 로봇은 인간이 만든 알고리즘을 가진 기계일 뿐이지만 인간과 관계를 맺고 나면 새로운 개체로 전환된다. 그럼에도 우리는 인공지능 로봇을 도구적 존재로 타자화(他者化)한다. 이러한 사고는 인간을 정복하는 인공지능 로봇 문제로 이어진다. 개인은 외부의 영향을 받고 사회적 경험을 통해 사회화되는 존재이며 개인의 의식은 집합 의식의 일부이다[15]. 이렇듯 상호작용 의례 이론을 통해 개인은 상황을 만들고 구성하는 결정적 요인이 아닌 성분으로 존재한다는 것을 이해해야 한다.

개인은 상황이라는 결과물을 혼자서 만들어 지지 않고 상호작용을 통해 상황이 바뀐다. 상황이란 공동체가 만들어낸 규범 속에서 만들어진다. 결국 특정한 상황도 공동체가 만들어낸 규범의 지배를 받는다[18]. 이처럼 인간의 긍정적 에너지 형성을 위해 도덕적으로 '선한 인공지능 로봇'을 개발한다는 것은 인간 개인의 수준으로는 환원할 수 없다.

기계는 인간과의 관계 속에서 의미를 갖는다는 시몽동의 시각처럼 인공지능 로봇은 인간의 목적에 따라 용도를 갖고 만들어진 것이지만 그것을 사용하는 인간의 본질적인 반응에 따라 바뀐다[23].

따라서 선한 인공지능 로봇이 재닛처럼 공정성과 공

감을 기반한 '거리두는 조력자'의 의미를 가지고 상호작용하게 된다면 인간과 기술의 관계는 상보적이고 공생적으로 존재할 수 있을 것이다. 그것은 상호작용 의례가 개인의 행동, 생각, 상황을 바꾸고 에너지의 강도에 따라 개인을 또 다시 변화시키기 때문이다[21].

현재 인공지능 개발에 있어 기술은 이전보다 뛰어나고 더 완벽한 인공지능을 바라면서 동시에 인간처럼 행동하는 인공지능만을 개발하고자 한다. 그러나 기술에만 집중한 무분별한 기술 개발은 또 다른 인공지능 문제를 양산할 수 있다. 앞으로 인간 삶의 문제는 인공지능 문제와도 연결되어 나타날 것이다. 결국 상황에서 상황으로 상호작용할 때 발현되는 의례 이론을 통해 공동체에 따라 나타나는 유대감, 정서적 에너지 등을 고려하여 인공지능의 역할을 탐색해야 한다. 이전까지는 상호작용 이론이 인간에게만 적용됐지만 이번 논의를 통해 인공지능 문제에도 적용하여 유대, 정서적 에너지, 집단상징, 도덕적 감정이라는 네 가지 상호작용 의례 이론에 부합하는 재닛의 역할 타당한가를 살펴보았다. 선한 존재인 재닛의 역할은 '거리두는 조력자'로서 인간의 요구사항을 모두 들어주면서 정서적 에너지를 증진시키는데 좋은 역할을 하지만 도움만 줄 뿐 개인들의 사적 편견에 직접 관여하지는 않는다. 재닛은 관찰자의 입장에서 도움을 주는 것에 그치고 있는 것이다. 이러한 인공지능은 책임성과도 연결되어 책임소재를 따질 때 인간과 인공지능에게 타당하고 적절한 윤리지침을 만들 수 있을 것이라 예측한다. 콜린스의 상호작용 이론을 통해 설명한 선한 인공지능에 대한 개념은 앞으로 인공지능 개발 패러다임을 인류와의 공진화에 맞출 때 기술개발이 지향할 수 있는 단서가 될 수 있을 것으로 기대한다.

## 참 고 문 헌

- [1] 김상태, "자율주행자동차에 관한 법적 문제," 경제규제와 법, 서울대학교 법학연구소, 제9권, 제2호, pp.177-190, 2016.
- [2] 이증기, "자율주행자동차 운행의 법적 이슈," 한국교통연구원, 교통연구, 제23권, 제3호, pp.2-4, 2016.
- [3] <http://www.hankookilbo.com/News/Read/2020>

02031544762632

[4] 김환석, *인공지능 시대를 보는 이론적 관점들*, 사회와 이론, pp.41-62, 2017.

[5] 손승혜, “우리는 어떤 로봇을 만들어가고 있는가? 2010-2019 텔레비전 뉴스에 등장한 인공지능 로봇의 전달 방식과 특성에 관한 탐색적 연구,” *한국언론학보*, 제64권, 제6호, pp.202-237, 2020.

[6] [http://www.chosun.com/site/data/html\\_dir/2018/10/11/2018101101250.html](http://www.chosun.com/site/data/html_dir/2018/10/11/2018101101250.html)

[7] <http://www.futuretimes.co.kr/news/articleView.html?idxno=13443>

[8] M. Tegmark, *Life 3.0*, 동아시아, pp.148-154, 2017.

[9] M. J. Sandel, *완벽에 대한 반론*, 와이즈베리, pp.112-113, 2016.

[10] A. Purington, G. J. Taft, S. Sannon, N. N Bazarova, and H. S. Taylor, “Alexa is my new BFF,” *Social Roles, User Satisfaction, and Personification of the Amazon Echo*, In *Proceedings of the 2017 CHI Conference Expanded Abstracts on Human Factors in Computing Systems*, pp.2853-2859, 2017.

[11] 김소연, “포스트휴머니즘 영화에서 (탈) 육체성과 기술-환상의 문제설정:〈블레이드 러너 2049〉를 중심으로,” *씨네포럼*, 제33호, pp.9-32, 2019.

[12] 이창섭, 이현정, “인공지능 혁신에 대한 기대와 불안 요인 및 영향 연구,” *한국콘텐츠학회논문지*, 제19권, 9호, pp.37-46, 2019.

[13] M. Scheler, *동감의 본질과 형태들*, 아카넷, 2006.

[14] 소병일, “공감과 공감의 윤리적 확장에 관하여: 흄과 막스 셸러를 중심으로,” *한국철학회*, 제118호, pp.197-225, 2014.

[15] R. Collins, *사회학 본능*, 알마, pp.244-250, 2014.

[16] C. Z. Elgin, “Fiction as thought experiment,” *Perspectives on Science*, Vol.22, No.2, pp.221-241, 2014.

[17] B. Reeves and C. L. Nass, *The media equation: How people treat computers, television, and new media like real people and places*, Cambridge university press, 1996.

[18] R. Collins, *사회적 삶의 에너지*, 한울아카데미, 2009.

[19] 신중섭, “도덕 감정과 이기심: 아담스미스를 중심으

로,” *철학논총*, 제73권, 3호, pp.109-133, 2013.

[20] G. H. Mead, *정신, 자아, 사회*, 한길사, pp.254, 2010.

[21] D. Boyns and S. Luery, “Negative emotional energy: A theory of the ‘dark-side’ of interaction ritual chains,” *Social Sciences*, Vol.4, No.1, pp.148-170, 2015.

[22] 김재희, “질베르 시몽동에서 기술과 정치,” *철학연구*, 철학연구회, 제108호, pp.123-151, 2015.

[23] 황수영, *질베르 시몽동*, 커뮤니케이션북스, pp.60-63, 2018.

저 자 소 개

추 미 선(Mi-Sun Chu)

정회원



- 2014년 2월 : 강원대학교 영상문화학과(문학사)
- 2019년 2월 : 강원대학교 영상문화학과 대학원(문학석사)
- 2021년 2월 : 강원대학교 영상문화학과 박사수료

<관심분야> : 인공지능 로봇, 드라마분석

유 승 호(Seoung-Ho Ryu)

정회원



- 1987년 2월 : 고려대학교 사회학과(학사)
- 1996년 2월 : 고려대학교 사회학과 대학원(사회학 박사)
- 2004년 ~ 현재 : 강원대학교 영상문화학과 교수

<관심분야> : 문화산업, 문화정책, 문화도시, 컴퓨터게임, HCI