

한국형 생성 인공지능 리터러시 척도 개발 및 타당화

Development and Validation of a Korean Generative AI Literacy Scale

노환호 (Hwan-Ho Noh) 연세대학교 바른ICT연구소¹⁾
김현정 (Hyeonjeong Kim) 연세대학교 바른ICT연구소²⁾
김민진 (Minjin Kim) 연세대학교 바른ICT연구소³⁾

< 국문초록 >

리터러시는 작성된 문서나 가공된 정보를 사람들이 읽고 이해할 수 있는 능력으로 시작되었다. 이후 디지털 기술이 발전하면서 정보를 접하고 활용할 수 있는 분야가 넓어짐에 따라 디지털 리터러시로 확장되었다. 디지털 리터러시의 활용 분야와 목적에 따라 다양한 용어로 사용되고 있다. 본 연구에서는 인공지능 시대에 그 중요성이 점차 높아지고 있는 생성 인공지능을 대상으로 이용자의 리터러시 수준을 확인하고, 선행 연구에서 제안한 리터러시의 개념을 확장하며, 한국 이용자에게 적합한 도구를 개발하고자 연구를 수행했다. 먼저 탐색적 요인 분석 결과, 생성 인공지능 리터러시는 AI 활용 능력, 비판적 평가, 윤리적 사용, 창의적 활용의 네 가지 요인으로 구성됨을 확인하였다. 다음으로 확인적 요인 분석을 통해 생성 인공지능 리터러시를 구성하는 네 가지 요인의 모형 구조가 통계적으로 적합하다는 것을 확인했다. 더불어 기존 인공지능 리터러시 척도와 인공지능 서비스 관련 평가 도구와의 상관 분석을 통해 새로 개발된 리터러시 도구가 실제 사람들의 태도와 유의한 관계가 있음을 확인하였고, 이를 통해 타당성이 확보되었음을 알 수 있었다. 마지막으로 본 연구의 의의와 한계점, 그리고 향후 연구 방향에 대해 논했다.

주제어: 생성 인공지능 리터러시, 측정 도구 개발, 타당성 평가, 윤리, 정보기술활용

1) 제1저자, hh_roh@naver.com

2) 교신저자, hjkim@barunict.kr

3) 공동저자, minjin.kim@barunict.kr

1. 서론

21세기 들어 인공지능에 대한 논의가 활발히 이루어졌으나 주로 기술에 대해 다루거나 전문가 영역에서 논의가 오가며 일반 이용자들의 현실 문제로 다가가지는 못했다(Ashok et al., 2022). 그러나 2022년 11월 생성 인공지능(Generative AI)인 ChatGPT가 등장한 이후 인공지능 시장은 더욱 가파른 속도로 성장하고 있다(OpenAI, 2024). 생성 인공지능은 인지적 처리 능력을 바탕으로 마치 사람처럼 기존 정보와 자료를 학습하고 학습 내용을 바탕으로 자연어 기반의 새로운 정보를 생산하는 기술로 정의된다(Wang, 2019). 이 과정에서는 머신러닝, 자연어 처리, 네트워크 기반 시스템을 주로 활용하여, 주어진 정보를 해석하고 처리한 후 새로운 정보를 생성한다(Zawacki-Richter et al., 2019). 인공지능이 주로 기술 개발이나 학습, 프로그래밍의 관점에서 논의가 이루어졌기에 생성 인공지능 또한 같은 관점에서 기술적인 측면이 강조되었다(Ashok et al., 2022).

그러나 현시대의 인공지능은 다양한 분야에서 사람들의 삶에 영향을 미치고 있으므로 전문가나 개발자의 관점에서 기술에 대해 이루어지는 논의가 아니라 현실에서 일반 사람들에게 미치는 영향과 활용 등의 관점에 대한 논의가 필요하다(Davenport & Ronanki, 2018). 이는 구글과 같은 정보 검색 서비스를 이용할 때도 인공지능 기반의 정보 탐색이 이루어지고, 유튜브나 소셜 미디어를 통한 게시물을 확인할 때도 자신의 관심 분야가 추천되는 등 다양한 방면으로 인공지능이 사람들의 선택에 영향을 미치고 있기 때문이다(Celik et al., 2022). 이외의 다양한 분야에도 인공지능이 도입되어 기존 시스템을 개선하고 효과성을 높이기 위한 시도가 이루어지고 있다. 그에 따라 디지털화로 인한 격차 문제가 인공지능 시대에도 반복될 수 있다는 의견이 제기되고 있다(Fui-Hoon Nah et al., 2023). 따라서 인공지능 시대에 사람들이 갖

추어야 할 인공지능의 활용 방법, 인공지능으로 인해 나타난 변화에 대한 적응, 그리고 이를 목적에 맞게 활용하는 방법에 대한 논의가 필요하다.

인공지능에 대한 사회적 논의가 필요한 이유는 인공지능이 언제나 긍정적인 영향만을 미치지 않는 때문이기도 하다. 인공지능 기술이 발전하면서 그에 따른 윤리적 문제나 악용 사례가 함께 보고되고 있어 이를 우려하는 목소리도 커지고 있다(Celik, 2023). 이러한 관점은 생성 인공지능 이전부터 다양한 경로를 통해 나타났다. 예를 들어, 기존 데이터를 학습한 인공지능이 성별 또는 인종 차별적인 결론을 도출하거나, 의도하지 않은 방향으로 부정적인 선택을 내리거나, 편향된 결론을 도출하는 등 과거의 나쁜 사례가 반복되는 문제가 발생했다(Ghallab, 2019; Wang & Siau, 2019). 이러한 문제는 생성 인공지능이 등장한 이후에도 지속되었다. 생성 인공지능은 자연어를 기반으로 명령하고 결과물을 얻을 수 있으므로 더욱 쉽고 빠른 사용이 가능하다(OpenAI, 2024). 그러나 이러한 특징에도 불구하고 기존의 편향이나 차별, 또는 ‘탈옥’으로 불리는 악용 사례가 지속적으로 나타났다(Orru et al., 2023). 자연어를 통한 프롬프트로 원하는 사항은 인공지능에게 전달하고 결과를 얻을 수 있다는 점은 생성 인공지능이 프로그래밍이나 번역, 정보 가공, 분석 등 다양한 영역에서 빠르고 쉽게 일을 도와줄 수 있다는 장점이지만, 생성된 정보의 품질과 정확성에 대해서는 아직도 많은 개선이 필요하다(Lund et al., 2023). 이는 생성 인공지능을 이용하고 이를 학습하기 위해서는 단순히 사용 방법이나 능력을 키우는 것을 넘어 이를 사용하는 과정에서 발생하는 윤리적 문제나 사회적 규범을 고려하고, 생성된 정보에 대해서도 비판적인 시각과 생산된 정보에 대한 전반적인 이해가 필요하다는 것을 의미한다. 따라서 기술에 대한 전반적인 이해와 함께 이를 윤리적이고 도덕적으로 사용하는 방법과 그 이유에 대한 학습이 필요하다고 할 수 있다.

이처럼 생성 인공지능은 자연어를 통해 누구나 쉽게 사용할 수 있지만, 사용하기에 따라 윤리적 문제가 발생할 수 있으며 때에 따라서는 사실이 아닌 오정보(misinformation)가 생성되기도 한다(Chelli et al., 2024; Lo, 2023; Walter, 2024). 따라서 생성 인공지능을 이해하고 이를 활용하기 위해서는 단순히 활용 역량과 능력만을 갖추는 것이 아니라 이를 올바르게 사용하고 좋은 결과를 낼 수 있는 역량을 갖추어야 한다. 이를 위해 선행 연구에서는 생성 인공지능 활용 능력을 개념화하고 이를 측정하기 위한 시도가 이어졌다(Celik, 2023). 여기서 생성 인공지능을 활용하고 결과를 산출하는 능력 전반을 인공지능 리터러시라고 한다. 리터러시는 기본적으로 글을 읽고 쓸 수 있는 역량을 의미하는 초기 개념에서부터 디지털 시대에 다양한 매체를 통해 접하고 얻을 수 있는 정보를 활용하고, 생산하며, 이를 기반으로 새로운 지식을 만드는 포괄적인 개념을 의미한다(Perry, 2012). 기술이 발전하면서 정보를 접할 수 있는 경로나 매체가 다변화됨에 따라 다양한 영역에 대한 리터러시의 개념화와 측정을 위한 시도가 이루어졌다(Brown, 1998; Buckingham & Burn, 2007; Koc & Barut, 2016; List, 2019; Saranto & Hovenga, 2004). 그에 따라 디지털 리터러시 또한 인공지능 리터러시의 차원으로 확대하려는 시도가 이루어졌으나, 아직 인공지능 리터러시를 개념화하고 측정하기 위한 연구는 다소 부족한 상황이다(Davenport & Ronanki, 2018; Ghallab, 2019).

이와 관련해 최근 인공지능 리터러시를 정의하고 측정하기 위한 연구들이 보고된다. 리터러시 도구 개발 연구들은 초기 읽고 쓰는 능력의 의미하는 리터러시 개념과 디지털을 통해 전달되는 정보에 처리와 생산에 관련된 리터러시 개념을 확장해 인공지능 리터러시를 측정하고자 했다(Carolus et al., 2023; Ng et al., 2021; Wang et al., 2023). 새로운 기술의 등장은 그 기술을 이해하고 활용하는 역량에 주목하게 하므로 리터러시 또한 활용 역량이나 능력이

강조된다(Aydin, 2021). 그러나 최근 생성 인공지능은 단순히 활용 역량만을 강조하는 것만으로는 온전히 리터러시를 개념화하지 못한다(Wang et al., 2023). 다른 기술과 달리 생성 인공지능은 자연어를 통해 쉽고 빠르게 명령을 내릴 수 있으며 결과 또한 자연어로 얻을 수 있어 이를 악용하기가 매우 쉽기 때문이다(Orru et al., 2023). 이러한 특징으로 인해 인공지능 리터러시에 대해 윤리적인 측면을 강조한 리터러시 개발을 추진하는 연구들이 이어졌으나, 인공지능을 통한 창의적인 활용이나 비판적 평가와 관련된 내용을 포괄한 연구는 부족한 상황이다(Ghallab, 2019). 더불어 기존의 인공지능 리터러시는 인공지능 전반에 대해 강조했으나 실제 사람들이 가장 많이 사용하고 문제가 되는 생성 인공지능 리터러시에 대해서는 많은 연구가 이루어지지 못했다(Fui-Hoon Nah et al., 2023).

최근, 생성 인공지능 사용 과정에서 발생할 수 있는 윤리적 문제와 정보에 대한 비판적 평가의 중요성이 제기되면서, 리터러시 측정 도구도 그 방향으로 개발되고 있다(Ng et al., 2024). 그러나 이들 연구 또한 서양 문화권의 사람들만을 대상으로 도구가 개발되어 한국인을 대상으로 측정하기 위한 도구가 부족한 상황이다. 특정 문화권이나 언어에 맞는 도구를 사용하지 않을 때 발생할 수 있는 의미상의 문제로 인해 한국형 도구를 만들기 위한 연구들이 이어지고 있으며(이병관 등, 2019), 생성 인공지능에 대한 이용자들의 이용 경험을 토대로 한 타당화 연구 또한 이어질 필요가 있다. 국내에서 발표된 논문 중 디지털 역량을 측정하거나(황용석 등, 2022), 인공지능 리터러시를 측정하더라도 문항이 외국어로 되어있고 인공지능의 기술적인 이해가 강조되거나(류미영, 한선관, 2023), 리터러시와 관련된 프레임워크(최숙영, 2022)나 역할(지유첸 등, 2022)에 대한 연구가 주를 이룬다. 그러나 이러한 선행 연구에도 불구하고 한국 이용자를 대상으로 생성 인공지능 리터러시를 개념화하고 이를 측정하기 위한 노력은 다소 이루어지지 못하고 있다. 이에 본 연구

에서는 인공지능을 기반으로 정의된 리터러시 개념을 확인하고 생성 인공지능 이용자에게 요구되는 리터러시를 측정하는 한국형 측정도구를 개발하는 것을 목표로 했다.

2. 이론적 배경

2.1. 인공지능 시대의 리터러시

디지털 시대를 넘어 인공지능 시대로 나아가는 현시점에서 리터러시의 개념도 점차 확대되었다. 기존 리터러시는 주로 사람들의 언어나 글을 읽고 쓰는 수준의 이해 능력을 일컫었다(Perry, 2012). 이후 정보통신 기술이 발전하며 정보가 디지털 환경으로 이전되었고 리터러시 개념 또한 그에 맞추어 확장되었다(Buckingham & Burn, 2007). 예를 들어 정보 리터러시(Saranto & Hovenga, 2004), 디지털 리터러시(List, 2019), 게임 리터러시(Buckingham & Burn, 2007), 미디어 리터러시(Brown, 1998) 등을 들 수 있다. 이후 등장하는 기술이나 서비스는 주로 디지털에 기반해 이루어지므로 디지털 리터러시가 새로운 리터러시 개념의 기본적인 뼈대가 되었다. 또한 디지털 리터러시는 사회적, 경제적 차원에서 다양한 자원이나 기회가 디지털을 통해 제공되고 얻을 수 있게 되면서 더욱 중요한 위치를 차지했다(van Laar et al., 2017). 디지털 리터러시는 디지털 영역에 국한된 것이 아니라 금융, 건강, 과학 기술 등의 영역으로도 확장되었다(Carolus et al., 2023). 디지털 리터러시가 기존의 리터러시와 차별화되는 것은 정보의 생산과 전달이 양방향으로 이루어진다는 것이다. 기존 미디어가 주로 단방향 정보 전달이 주된 역할을 했다면, 뉴미디어를 비롯한 디지털 시대의 새로운 서비스는 사람들이 정보를 단순히 받아들이기만 할 뿐 아니라 새로운 정보를 생산하고 공유할 수 있도록 만들었다(Butler, 2012). 따라서 자신이 접한 정보에 대한

이해를 넘어 다른 사람이 만들어 공유한 정보를 비판적인 시각으로 바라보고, 그 정보를 자신이 다시 새로 가공해 새로운 정보를 만들어 낼 수 있는가와 같은 정보 생산자의 역할과 능력이 리터러시에 포함되었다(Koc & Barut, 2016). 이는 디지털 공간이 하나의 사회적 교류의 장이 되면서 사람들의 역할 또한 더욱 확장되었다고 볼 수 있다(Lin et al., 2013).

기술을 사용하기 위해서는 기술에 대한 기본적인 지식이나 사용 방법에 대한 이해가 필요하다(Aydin, 2021). 제품 설명서나 이용 가이드 등의 정보를 전달하는 것은 사람들이 서비스나 기술을 이해하고 원활히 사용하도록 하기 위함이다. 예를 들어 인터넷에 접속해 정보를 탐색할 때도 사람들은 인터넷에 대한 이해나 사용 방법, 경험이 많을수록 인터넷을 더욱 효과적으로 이용할 수 있다(Rajam et al., 2021). 이 과정에서 사람들은 인터넷에 대한 지식과 경험이 쌓이며 활용 능력이나 기술에 대한 이해를 포함해 전반적인 시스템에 대한 이해가 높아진다. 또한 기술이나 서비스를 이용하면서 그 기술의 강점이나 효과성을 깨닫게 되면, 이를 자신의 모든 일상에 적용해 리터러시를 비롯한 전반적인 역량도 높아진다(Wang & Wu, 2022). 이와 같은 전반적인 과정을 통해 특정 영역에 대한 개인의 리터러시가 높아진다. 리터러시를 높이기 위해서는 이와 같은 경험과 학습이 중요하므로 사람들 간의 기회의 부족이나 역량의 차이로 인해 디지털 격차가 발생하게 된다(van der Zeeuw et al., 2019). 따라서 그 원인은 다양할 수 있으나 리터러시가 낮다는 것은 결국 그 시대나 기술이 요구하는 역량과 그에 따른 혜택의 부족으로 인해 다양한 부분에서 약점이 될 수 있다는 것을 의미한다. 그러므로 디지털 시대를 넘어 인공지능 시대로 나아가는 현시점에서 인공지능 리터러시는 가장 시급하고 중요한 문제라고 할 수 있다(Celik, 2023). 즉, 인공지능을 이해하고, 활용하며, 인공지능을 통해 얻을 수 있는 혜택을 누리기 위해서는 인공지능에 대한 종합적인

지식을 의미하는 리터러시의 개념을 학습하고 이해할 필요가 있다(Acharya et al., 2023). 인공지능의 중요성이 점차 커지며 다양한 분야에 인공지능 기술이 적용되어 인공지능 리터러시가 필수적인 역량으로 주목받고 있기 때문이다(Ng et al., 2021).

생성 인공지능을 이해하기 위해서는 다양한 분야에 대한 이해도 필요하다. 생성 인공지능은 기존 컴퓨터나 정보과학의 영역을 넘어 심리학이나 사회학 등의 측면에서 생산된 정보가 미치는 영향이나 자연어를 통한 상호작용 방법에 관한 연구까지 다양한 차원에서 이루어지고 있다(Southworth et al., 2023). 자연어를 통해 직접 생성 인공지능과 상호작용하며 정보를 생산하고 그 정보를 이해하며 가공하는 것이 필요하므로, 사회적 상호작용이나 사회적 규범, 문화적 요인에 대한 고려도 필요하다(Poria et al., 2017). 따라서 인공지능 리터러시는 사용 능력이나 생성된 정보에 대한 평가를 넘어 기술에 대한 전반적인 이해와 윤리적 사용 측면을 함께 고려하는 방향으로 확장되었다(Wang et al., 2023). 생성 인공지능을 활용하기 위해서는 인공지능에 대한 전문적인 수준의 기술적 이해가 아니라 생성 인공지능과 사용자가 자연어를 통해 상호작용하면서 사회문화적으로 적합하고 올바른 방향으로 정보를 활용해야 한다(Robinson, 2020). 인공지능 리터러시는 점차 현대의 필수적인 역량으로 여겨지며 다양한 차원에서 리터러시를 높이기 위한 노력도 요구된다(Carolus et al., 2023).

인공지능 리터러시는 다음과 같은 차원에서도 중요하다. 생성 인공지능은 인공지능 기반의 학습 모델을 활용해 정확한 데이터 처리와 의사결정에 도움을 준다(Wang & Siau, 2019). 생성 인공지능은 기존 데이터를 기반으로 학습을 통해 현재의 시스템과 데이터에 최적화할 수 있는 규칙을 이해하고, 새로운 정보나 가치를 만들어 의사결정을 돕는다(Kong et al., 2022). 이와 같은 과정은 사람들의 인지적 처리 과정을 통해 문제 해결과 유사한 양상을 보인다. 사람들은 문제 해결을 위해 다양한 요소와 정

보를 바탕으로 문제를 이해하고, 새로운 문제에 관한 대응 방법이나 문제 해결에 관한 정보를 전달한다(Wang, 2019). 생성 인공지능을 이용하는 것은 문제 해결을 위해 고려해야 할 다양한 정보나 판단에 필요한 데이터, 기존에 문제 해결을 위해 적용되었던 규칙과 패턴에 대한 이해를 효과적이고 효율적인 방법으로 돕기 때문에 이를 활용하는 것이 인공지능 시대에 주요한 능력이라고 할 수 있다. 이와 관련된 선행 연구에서는 학생들의 인공지능 활용과 자신감의 수준을 측정하기 위한 도구를 만들기도 했으며(Chai et al., 2020), 인공지능 학습 동기를 측정하는 방향으로도 확장되었다(Lin et al., 2021).

인공지능 리터러시는 인공지능을 단순히 활용하는 것을 넘어 인공지능 서비스를 통해 정보를 생산하고 평가하는 능력을 요구한다(Ng et al., 2024). 기존 인공지능 리터러시는 주로 인공지능 서비스를 이용하고, 생산된 정보를 평가하고 이를 활용하는 측면을 주로 다뤘다(Ng et al., 2021). 그러나 최근 들어 인공지능의 영향력이 확대됨에 따라 인공지능을 통해 생산된 정보가 잘못 사용될 수 있다는 점을 고려해 비판적인 태도나 평가를 포함하기 위한 노력이 이루어지고 있다(Ng et al., 2024). 그에 따라 인공지능 사용으로 인해 발생할 수 있는 공정성, 설명 가능성, 투명성, 윤리성, 포용성 등의 문제들이 중요하게 다루어져야 한다는 목소리가 높아진다(Hagendorff, 2021). 더욱이 생성 인공지능이 자연어를 통해 쉽게 사용할 수 있음에 따라 이와 같은 사회적, 윤리적, 도덕적 차원의 인식이 더욱 중요하다고 할 수 있다(Southworth et al., 2023). 이를 고려해 인공지능 리터러시 측정도구를 개발한 선행 연구에서는 인공지능 인지, 활용, 평가 그리고 윤리의 관점에서 인공지능 리터러시 구성 개념을 제안했다(Wang et al., 2023). 또 다른 연구에서는 인공지능 리터러시를 활용과 적용, 이해, 탐지 및 윤리의 관점으로 제시했다(Carolus et al., 2023). 최근 Ng et al.(2024)의 연구에서는 인공지능에 대한 행동적, 정서적 평가만이 아

나라 인지적, 윤리적 측면에 대해 고려한 인공지능 리터러시 도구를 개발하고자 했다. 그러나 특정 연령대를 위한 리터러시 개발에 초점을 맞추었으며 주로 교육 측면에서 인공지능 리터러시를 학습하고 이해하는 측면을 강조하고 있다. 따라서 다양한 생성 인공지능에 대한 사람들의 전반적인 리터러시 수준을 측정하기 위해서는 다양한 사람들을 대상으로 일반적인 영역의 리터러시 수준을 개념화하고 측정하는 것이 필요하다고 할 수 있다.

2.2. 생성 인공지능 리터러시 구성 개념

디지털 공간에서 정보의 홍수라고 일컬을 만큼 많은 정보가 유통되면서 그 정보를 명확히 이해하고 활용할 수 있는 역량에 대한 중요성이 강조된다. 이와 같은 역량을 뉴미디어 리터러시라고 하며 주어진 정보에 대한 사실 확인을 넘어 정보를 생산하고 소통할 수 있는 역량이 포함된다(Koc & Barut, 2016). 기존의 미디어는 주로 주어진 정보에 대한 이해와 학습의 영역을 강조했으나, 뉴미디어 리터러시는 정보 생산, 비판적 평가 그리고 공유를 강조한다(Ritzer & Jurgenson, 2010). 정보에 대한 이해와 판단에 있어서 가장 중요한 부분은 정보를 비판적으로 소비하고 생산하는 것이기 때문이다(Chen et al., 2011). 정보에 대한 사실관계를 확인하는 것은 주어진 정보를 단순히 받아들이기만 하는 것이 아니라 그 정보를 기반으로 새로운 정보를 만들거나, 사실관계를 확인할 수 있는 단서를 찾고, 이를 검토하고 처리하는 복합적인 과정이 동반된다(Brashier & Marsh, 2020).

OpenAI의 ChatGPT 등장과 생성 인공지능 기술 발전은 이제 인공지능이 학습하고 생성한 정보에 대한 이해와 신뢰에 대한 영역에서도 사람들의 이해 역량이 중요하게 작용한다(Dunn et al., 2023). 온라인에서 활발히 활동하는 사람일수록 더 많은 정보를 접하게 될 가능성이 높으며 그에 따라 정보에 대한 사실관계를 확인해야 하

는 경우도 더욱 많아진다(Livingstone & Helsper, 2010). 이는 디지털 이용 역량이 높은 학생일수록 온라인 공간에서의 유해 콘텐츠가 노출될 가능성이 높아지는 것과는 관련된다(Staksrud et al., 2013). 더욱이 악의적인 목적으로 만들어진 오정보(misinformation)와 관련된 문제는 현대 사회에서 중요한 문제가 되고 있으며(Pennycook et al., 2020), 그 사실관계를 확인하기가 매우 어렵다는 특징이 있다(Tan et al., 2015). 잘못된 정보가 사람들에게 문제가 되는 이유는 잘못된 정보가 미치는 영향이 자신보다 타인에게 더욱 심각한 문제를 유발한다는 과대평가(Jang & Kim, 2018)와 함께 그 정보가 자신이 사실 알고 있던 사실이거나 틀리지 않다고 생각하는 인지적 편향을 유발하기 때문이다(Allcott & Gentzkow, 2017). 잘못된 정보가 무분별하게 확산한다는 인식은 사람들에게 온라인 공간에서 접하는 정보에 대한 위험성 인식을 높이며, 자신이 정보를 얼마나 잘 평가할 수 있을 것인지에 대한 개인의 자신감 수준이 정보 위험성 인식에 주된 영향을 미친다(Duffy, 2018). 이는 사람들이 동기화된 추론을 통해 주어진 상황을 이해하고 판단하기 때문이다(Kruger & Dunning, 1999). 따라서 윤리적인 차원에서 정보를 편향되거나 차별적으로 다루지 않는 것이 중요하다.

사람들은 정보를 통해 다른 사람과 소통하고 상호작용하려 하므로, 디지털 환경에서는 주어진 정보를 이해하는 것뿐만 아니라, 비판적으로 판단하고 생성하거나 공유할 수 있는 역량이 중요하다(Koc & Barut, 2016). 사람들은 불확실한 상황을 피하고 주어진 현상을 이해하려는 기본적인 동기가 있다(Namkoong & Henderson, 2016). 오정보가 만연한 사회일수록 이를 대처하고 손실을 미리 방지하려는 경향이 높아지므로(Simonsohn, 2009), 리터러시 수준은 정보에 대한 이해와 판단을 결정하는 주된 도구가 될 수 있다. 선행 연구에 따르면 사람들은 인지적 처리 능력이 높고, 심사숙고한 사고방식을 사용하고, 관련된 지식수준이 높을수록 정보의 사실관계를 파악할

가능성이 높다(Bago et al., 2020; Bronstein et al., 2019; Pennycook & Rand, 2019). 따라서 리터러시 수준이 높은 사람일수록 주어진 정보에 대한 이해와 사실관계의 확인이 더욱 분명하고 명확하게 나타날 것이다. 생성 인공지능을 이용하는 상황에서 이용자의 정보 이해와 비판적 평가 영역을 함께 살피는 것은 단순히 주어진 상황에서 정보 이해 역량을 넘어 실제적인 정보의 활용과 이용 능력 평가를 반영하는 리터러시 개념을 보여줄 수 있을 것이다. 따라서 생성 인공지능을 활용하는 차원에서도 생성된 정보에 대한 비판적 평가가 중요하다.

또 다른 차원에서는 인공지능을 창의적으로 활용하는 방안에 대해 고려할 수 있다. 생성 인공지능을 이용하기 위해서는 전문가 수준의 코딩 역량이나 상호작용 능력을 요구하는 것이 아니라 일상적인 수준의 단어와 문장을 사용하는 것이 필요하다(Cain, 2024). 이 과정에서 생성 인공지능에게 어떤 명령어를 제시하는지에 따라 그 결과물은 확연한 차이가 발생할 수 있다(Jiang et al., 2023). 따라서 전문적인 수준의 기술을 다루는 사람이 아닌 일반적인 이용자에게도 생성 인공지능을 통해 정보를 생산하는 과정에서 이를 정보에 대한 비판적인 평가뿐만 아니라 효과적이고 창의적으로 사용하는 능력이 요구된다(Knoth et al., 2024). 생성 인공지능이 등장한 이후 인공지능 기술은 더 이상 미래 기술이 아닌 현재의 문제가 되었으며, 이를 얼마나 잘 활용할 수 있는지가 중요하다(Fui-Hoon Nah et al., 2023). 효과적인 정보 생산을 위해서는 문제를 해결할 수 있는 창의적인 방법에 대한 접근이 필요하며, 이와 같은 역량을 키워야 한다(Lo, 2023). 생성 인공지능이 만드는 정보는 입력하는 프롬프트가 얼마나 명확하고 창의적인가에 달려 있으므로, 이를 적절히 활용할 수 있는 역량을 갖추는 것이 필요하다. 학습적 차원에서도 점차 생성 인공지능을 얼마나 잘 활용할 수 있는가에 따라 얻을 수 있는 결과가 달라지는 만큼 교육 과정에서도 창의적인 활용을 강조한다(Walter, 2024). 이

는 생성 인공지능을 효과적으로 사용하기 위해 창의적 활용이 중요하다는 것을 의미한다.

이에 본 연구에서는 인공지능 산업이 다양한 영역으로 확장되고, 다양한 계층에서 이를 활용하는 역량을 요구하고 있는 만큼 본 연구에서는 전 연령을 대상으로 일반적인 수준의 생성 인공지능 리터러시 측정 도구를 개발하고자 했다. 또한 현재 서비스되는 주요 앱이나 프로그램에서 사용되는 생성 인공지능에 초점을 맞추어 정보 생산과 활용 그리고 이에 대한 비판적인 태도를 강조하는 도구를 개발해 생성 인공지능 서비스에 대한 이용자의 리터러시를 측정하는 도구를 개발하는 것에 목적을 두었다. 이를 위해 리터러시 개념을 네 가지 영역으로 구분하였으며 선행 연구에서 확인한 세 가지 영역에 윤리적인 부분을 함께 다루고자 했다(Ng et al., 2021). 인공지능에 대한 이해와 사용 능력을 다룬 ‘AI 활용능력’을 기반으로 새로운 개념이나 정보를 생산하고 문제를 해결하는 ‘창의적 활용’, 정보에 대한 평가와 비판적 시각을 다룬 ‘비판적 평가’ 그리고 인공지능 사용 과정에서 발생하는 윤리적 문제를 고려한 ‘윤리적 사용’의 네 가지 요인을 구성하기 위한 측정도구 개발 연구를 수행했다.

3. 연구 방법 및 결과

3.1. 한국형 생성 인공지능 리터러시 척도 개발 필요성

한국형 생성 인공지능 리터러시 척도를 개발하는 이유는 측정 도구가 전달하고자 하는 내용이나 그 맥락이 도구에 충분히 반영되지 않을 경우, 측정하고자 하는 개념을 정확히 전달하는 데 한계가 있기 때문이다. 단순히 문장을 번역하는 수준에서는 척도가 전달하고자 하는 의미가 왜곡될 가능성이 높으며, 이는 연구의 신뢰성과 타당성에 부정적인 영향을 미칠 수 있다(김세나 등, 2015). 따라서 한국어

사용자에게 개념을 온전히 전달하고, 그 의미가 명확하게 이해되었는지를 확인할 수 있는 도구가 필요하다.

기존 연구에서도 한국적 맥락을 반영한 척도 개발의 중요성이 강조되었다. 서양 중심의 척도가 한국적 환경에서 정확한 의미 전달이 어려울 수 있으며, 이를 해결하기 위해 한국적 특수성을 반영한 새로운 척도를 개발하는 것이 필요하다(이병관 등, 2019). 따라서 본 연구에서는 한국적 맥락에서 적합한 생성 인공지능 리터러시 척도를 개발하고자 하며, 이를 통해 한국어 사용자에게 AI 리터러시 개념을 정확히 전달하고, 해당 개념을 명확하게 측정할 수 있는 도구를 제공하고자 한다. 이는 연구의 신뢰성과 타당성을 높이고, AI 관련 교육과 정책 개발에 실질적인 기여할 수 있을 것이다.

특히 한국 사회에서 AI 기술이 급속도로 발전하고 있는 점을 고려할 때, 기존 서구 중심의 리터러시 개념만으로는 충분히 대응할 수 없는 측면이 존재한다. 한국은 세계적인 디지털 기반을 갖춘 사회로, 인공지능 기술의 적용이 매우 활발한 국가이다. 이에 따라 AI 기술을 활용한 창의적인 문제 해결 및 윤리적 사용에 관한 관심이 높아지고 있다(방송통신위원회, 2024). 한국 사회는 단순히 기술을 사용하는 것에 그치지 않고, AI를 통해 창의적인 가치를 창출하려는 경향이 강하며(Southworth et al., 2023), 이는 본 연구에서 ‘창의적 활용’이라는 차원을 추가한 이유이기도 하다. 또한 생성 AI 사용 관련 윤리적 이슈가 중요한 사회적 주제로 떠오르면서(Lund et al., 2023), 한국적 맥락에 맞는 윤리적 사용 기준이 척도에 포함될 필요가 있다.

3.2. 척도 개발 절차

본 연구에서는 선행 연구 리뷰를 통해 선정 한 네 가지 요인(AI 활용능력, 창의적 활용, 비판적 평가, 윤리적 사용)을 측정하기 위해 인공지능 리터러시 측정 도구의 문항을 개발했다. 이 과정에서 ICT 관련 연구소에 재직 중

인 박사급 연구원 3인이 참여했으며, 문항 개발은 DeVellis(2017)에서 제안한 심리적 측정 도구 개발 절차를 따라 진행되었다. 또한, 본 연구는 개인의 주관적인 역량 수준을 평가하기 위한 것이므로, 관련된 선행 연구의 개발 지침과 과정을 참고했다.

우선, 문항 개발을 위해 인공지능 리터러시 선행 연구를 바탕으로 네 가지 요인의 구성 개념을 정의하고, 선행 연구에서 제시된 측정 지표를 정리했다. 이후, 선행 연구의 문항을 한국어로 번안한 후, 생성 인공지능 리터러시 측정 도구에 부합하도록 수정과 도구 개발 과정을 거쳤다. 이 과정에서 Wang et al.(2023), Ng et al.(2021), Carolus et al.(2023)의 연구에서 제안된 인공지능 리터러시 측정 도구를 참고하여 문항을 개발했다.

먼저, Wang et al.(2023)의 연구에서는 인공지능 리터러시를 ‘인지’, ‘활용’, ‘평가’, ‘윤리’라는 네 가지 주요 요소로 구분하여 측정 도구를 개발했다. 이 연구는 디지털 리터러시의 개념을 기반으로 하며, 인공지능 활용에 대한 인지적 요소와 윤리적 고려 사항을 중점적으로 다뤘다. 본 연구에서도 이러한 요소를 중요하게 고려하여, 특히 ‘윤리적 사용’과 ‘비판적 평가’라는 요인에서 Wang 등의 연구 문항을 한국어로 번안하고 수정했다. 예를 들어, 사용자가 AI를 어떻게 인식하고 평가하는지를 측정하는 문항들은 Wang et al.(2023)의 연구에서 제안된 요소를 기반으로 개발했다.

다음으로, Ng et al.(2021)의 연구는 인공지능 리터러시의 개념을 ‘AI 이해’, ‘AI 활용’, ‘AI 평가 및 생성’, ‘AI 윤리’라는 네 가지 범주로 분류했다. 이 연구는 AI 리터러시의 다양한 측면을 다루며, 특히 사용자의 AI 이해와 활용 능력에 초점을 맞추고 있다. 본 연구에서는 Ng 등의 연구를 바탕으로 ‘AI 활용능력’과 ‘창의적 활용’ 요인에 대한 문항을 개발했다. 예를 들어, AI 기술의 기본 개념을 이해하고 이를 다양한 상황에서 적용할 수 있는 능력을 측정하는 문항을 개발할 때, Ng et al.(2021)의 연구에서 제시된 문항을 번안하고 수정하여 한국적 맥락에 맞게 적용했다.

마지막으로, Carolus et al.(2023)의 연구에서는 AI 리터러시 척도를 개발할 때 심리적 역량과 메타 역량을 포함하는 접근을 취했다. 이 연구는 사용자가 AI와 상호작용하는 과정에서 문제 해결 능력, 학습 능력, 감정 조절 능력 등 심리적 역량이 중요함을 강조했다. 본 연구에서는 Carolus et al.(2023)의 연구에서 제안된 ‘AI 문제 해결’, ‘AI 학습’, ‘AI 감정 조절’의 요소를 참고하여 한국어 문항을 개발하는데 적용했다. 예를 들어, 사용자가 AI를 사용하면서 비판적 태도와 평가를 측정하는 문항은 Carolus et al.(2023)이 제안한 메타 역량의 개념을 기반으로 하여 개발했다.

이와 같이, Wang et al.(2023), Ng et al.(2021), Carolus et al.(2023)의 연구에서 제안된 문항과 개념을 바탕으로 본 연구의 척도 개발을 진행했으며, 이를 통해 한국적 맥

락에서 타당하고 신뢰할 수 있는 생성 인공지능 리터러시 척도를 개발하고자 했다. 초기에는 32개의 문항이 개발되었고, 네 가지 요인을 측정하는 문항으로 각 8개씩 구성되었다. 또한, 한국형 생성 AI 리터러시 척도를 개발하기 위해 기존 영문 문항을 번안하는 과정에서, 번안된 문항의 타당성은 영어 사용에 능통한 해외 대학 출신 박사 연구원의 도움을 받아 진행했다.

3.3. 탐색적 요인 분석(exploratory factor analysis)

3.3.1. 연구 방법 및 절차

생성 인공지능 척도의 일반화를 위해 20-60대의 다양

〈표 1〉 탐색적 요인 분석 연구 참가자 인구통계 정보

| 특성 | | 빈도(명) | 비율(%) |
|--------------------|--------------|-------|-------|
| 성별 | 남성 | 500 | 50 |
| | 여성 | 500 | 50 |
| 연령 | 20대 | 200 | 20 |
| | 30대 | 200 | 20 |
| | 40대 | 200 | 20 |
| | 50대 | 200 | 20 |
| | 60대 | 200 | 20 |
| | | | |
| 생성 AI 서비스 이용 경험 | 예 | 731 | 73.1 |
| | 아니오 | 269 | 26.9 |
| 최근 생성 AI 서비스 이용 시기 | 최근 한 달 이내 | 391 | 39.1 |
| | 이용한 지 한 달 넘음 | 261 | 26.1 |
| | 한 번만 사용해 보았다 | 79 | 7.9 |
| | 이용 경험 없음 | 269 | 26.9 |
| 학력 | 고졸 이하 | 117 | 11.7 |
| | 대학 재학 또는 졸업 | 725 | 72.5 |
| | 대학원 재학 또는 졸업 | 158 | 15.8 |
| 주관적 사회경제적지위 | 최하 | 12 | 1.2 |
| | 하 | 86 | 8.6 |
| | 중하 | 253 | 25.3 |
| | 중 | 373 | 37.3 |
| | 중상 | 244 | 24.4 |
| | 상 | 31 | 3.1 |
| | 최상 | 1 | 0.1 |
| Total | | 1,000 | 100 |

한 계층을 대상으로 한 도구 개발을 위해 온라인을 통한 설문조사를 진행했다. 온라인 조사회사 패널 중 생성 인공지능에 대해 알고 있는 사람을 설문조사 참가자로 선정했다. 총 1,000명의 대한민국 성인 남녀를 모집했으며, 세대별 200명의 참가자를 남녀 균등 모집했다. 탐색적 요인 분석을 위해서는 최소 200명 이상의 표본이 필요하고(MacCallum et al., 1999), 개발된 도구 문항 수의 최소 5배 이상을 모집하는 것을 권장한다(Hair et al., 2010). 따라서 본 연구에 참여한 표본 수는 탐색적 요인 분석 수행에 충분한 수가 확보되었다고 할 수 있다. 설문에 참여한 사람들의 특징을 확인할 수 있도록 인구통계 문항과 함께 본 연구에서 개발한 32개의 초기 문항이 설문에 포함되었다. 참가자들은 설문 참여에 동의한 뒤 주어진 설문 문항에 답했다.

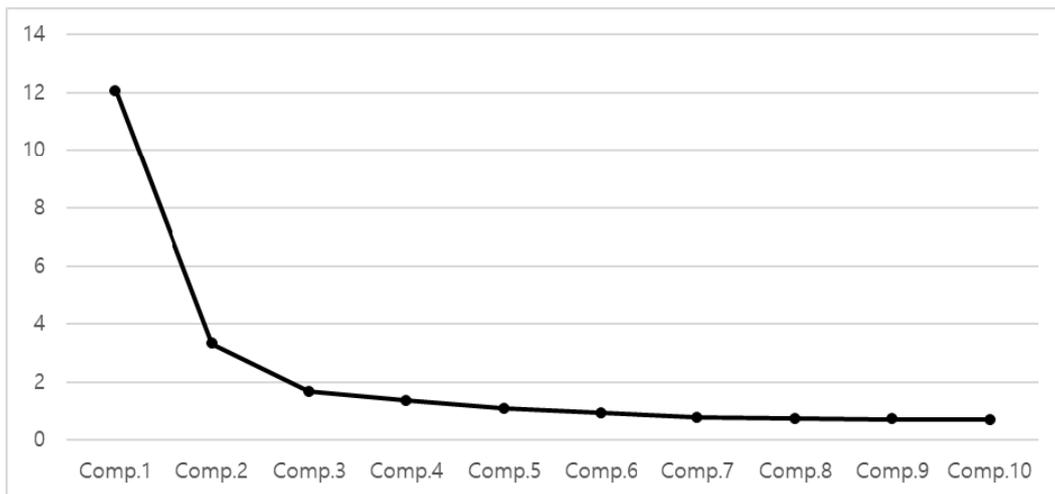
3.3.2. 연구 결과

온라인을 통해 수집된 데이터를 기반으로 탐색적 요인 분석(Exploratory Factor Analysis; EFA)을 수행했다. 탐색적 요인 분석은 R 프로그램(R Development Core Team, 2012)을 사용했으며 psych 패키지를 통해 분석을 수행했다(Revelle, 2018). 탐색적 요인 분석을 위한 요인

추출은 최대우도법(Maximum likelihood)을, 요인 회전은 요인 간 상관을 가정한 사각 회전 방법 중 가장 대표적인 Promax 방식을 사용했다(Hetzl, 1996). 또한 측정 도구의 신뢰도 검증을 위해 지표의 내적 타당성을 검증하는 Cronbach's alpha 수치를 확인했다(.70 이상; Nunally, 1978).

우선 본 연구에서 가정한 네 요인이 적합한지 확인하기 위해 탐색적 요인 분석을 수행하기 전 요인 수를 결정하기 위해 스크리 도표(scree plot) 확인 과정을 거쳤다. 스크리 도표는 현재 구성된 문항을 통해 도출할 수 있는 요인에 대한 주성분 분석을 통해 고유값 1을 기준으로 요인 수를 결정하는 도표이다. 이 도표에서 고유값의 기울기가 완만해지는 시점을 기준으로 요인 수를 선정할 수 있다. 그래프를 확인한 결과 본 연구에서는 네 요인 모형이 타당한 것으로 나타났다 <그림 1>. 따라서 네 요인 모형을 기반으로 탐색적 요인 분석을 수행했다.

이어서 개발된 전체 문항을 대상으로 탐색적 요인 분석을 수행하여 여러 요인에 걸쳐서 요인 부하량(factor loadings)이 걸리는 문항과 요인 부하량의 수치가 낮은 문항(.40 미만; Nimon et al., 2011) 등을 제거하는 과정을 수행했다. 문항 제거는 요인 부하량 수치뿐만 아니



<그림 1> 탐색적 요인 분석 데이터 기반 스크리 도표 분석 결과

라 공통 분산(communality)이 높으며(.50 이상), 요인 부하량이 여러 요인에 걸치는 수준을 의미하는 복잡성(complexity for an item)이 낮은(2.0 이하) 문항을 선별하고자 했다. 또한 문항 선별과 제거를 위해 문항을 실제로 확인하고 문항을 제거하는 것이 타당한지에 대해서 연구진의 안면 타당성(face validity) 검증을 추가로 진행했다. 그러나 공통 분산이 낮더라도 본 연구의 하위 요인을 구성하기에 적합하여 연구진의 논의를 통해 안면 타당도 검증을 거친 문항을 추가해 최종적으로 12개의 문항이 최종 문항으로 선정되었다. 선정된 문항은 <표 2>를 통해 확인할 수 있다.

선별 과정을 거쳐 남은 문항을 대상으로 분석 결과 네 요인 모형은 전체 설명량 중 61%를 설명했다. 네 요인 모형의 모형 적합도는 Tucker-Lewis Index(TLI; Bentler & Bonett, 1980)가 .993, RMSEA(Steiger, 1990)는 .023 [90% CI .000, .053]으로 분석을 수행하기에 적합하다는 것을 확인할 수 있었다(Kline, 2015). 측정 도구의 요인

부하량은 모두 .50 이상으로 확인되었다. 스크리 도표와 고유값을 기준으로 할 때 다섯 요인 또한 분석을 수행할 수 있는 조건으로 보여 다섯 요인에 대해서도 추가로 분석했다. 추가 분석 결과 모형 적합도는 적합한 수준으로 나타났으나, 추가된 요인의 설명량이 5% 이하이거나 다른 요인과 상관관계가 낮게 나타나 생성 인공지능 리터러시를 구성하기에 적합한 요인이라고 판단되지 않았다. 따라서 생성 인공지능 리터러시 척도는 최종적으로 네 요인 모형을 선정했다. 이 과정을 통해 초기 32개 문항에서 12개 문항의 최종 선정되었다.

3.4. 확인적 요인 분석(confirmatory factor analysis)

3.4.1. 연구방법 및 절차

개발된 12문항, 네 요인 모델의 통계적 모형 적합성을 확인하기 위해 확인적 요인 분석을 수행했다. 확인

<표 2> 생성 인공지능 리터러시 척도 탐색적 요인 분석 결과

| 문항 | AI 활용능력 | 비판적 평가 | 윤리적 사용 | 창의적 활용 |
|--|------------|-----------|-----------|-----------|
| 나는 인공지능 챗봇을 사용해 필요한 정보를 얻을 수 있다. | 0.86 | 0.01 | 0.01 | -0.07 |
| 나는 인공지능 챗봇과 대화를 통해 원하는 정보를 얻을 수 있다. | 0.73 | -0.09 | 0 | 0.05 |
| 나는 인공지능 챗봇과 상호작용해 원하는 작업을 수행할 수 있다. | 0.71 | 0.07 | 0 | 0.03 |
| 나는 인공지능 챗봇 활용에 따른 윤리적 및 사회적 영향을 비판적으로 평가할 수 있다. | -0.04 | 0.93 | -0.03 | -0.09 |
| 나는 인공지능 챗봇의 긍정적인 효과와 부정적인 효과를 비판적으로 이해하고 평가할 수 있다. | 0.04 | 0.80 | 0 | -0.04 |
| 나는 인공지능 챗봇이 생성한 정보에서 사실과 의견을 구분할 수 있다. | -0.03 | 0.53 | 0.1 | 0.17 |
| 나는 법적/윤리적 관점에서 인공지능 챗봇이 제공하는 정보를 평가한다. | -0.08 | -0.06 | 0.86 | 0.10 |
| 나는 인공지능 챗봇의 사용과 관련된 윤리적 책임을 이해하고 실천할 수 있다. | 0.11 | 0 | 0.80 | -0.02 |
| 나는 인공지능 챗봇 사용 시 개인정보와 정보 보안에 대해 항상 경계한다. | -0.02 | 0.04 | 0.58 | -0.08 |
| 나는 인공지능 챗봇을 활용해 창의적인 문제 해결이 가능하다. | -0.09 | -0.09 | -0.01 | 0.97 |
| 나는 인공지능 챗봇을 사용해 복잡한 문제를 해결할 수 있다. | 0.15 | 0 | -0.02 | 0.65 |
| 나는 인공지능 챗봇을 통해 새로운 관점에서 문제에 접근할 수 있다. | 0.11 | 0.13 | 0.01 | 0.56 |
| Eigenvalues | 1.88 | 1.8 | 1.75 | 1.74 |
| % of Variance | 0.16 | 0.15 | 0.15 | 0.15 |
| Proportion Explained | 0.26 | 0.25 | 0.24 | 0.24 |
| Cronbach's alpha | 0.813 | 0.805 | 0.749 | 0.806 |

〈표 3〉 확인적 요인 분석 연구 참가자 인구통계 정보

| 특성 | | 빈도(명) | 비율(%) |
|--------------------|--------------|-------|-------|
| 성별 | 남성 | 237 | 47.4 |
| | 여성 | 263 | 52.6 |
| 연령 | 20대 | 50 | 10 |
| | 30대 | 120 | 24 |
| | 40대 | 203 | 40.6 |
| | 50대 | 105 | 21 |
| | 60대 | 22 | 4.4 |
| 생성 AI 서비스 이용 시작 시기 | 2022 | 65 | 13 |
| | 2023 | 318 | 63.6 |
| | 2024 | 117 | 23.4 |
| 최근 생성 AI 서비스 이용 시기 | 매일 사용함 | 160 | 32 |
| | 일주일 이내 사용하였음 | 252 | 50.4 |
| | 한 달 이내 사용하였음 | 88 | 17.6 |
| 학력 | 고졸 이하 | 65 | 13 |
| | 대학 재학 또는 졸업 | 379 | 75.8 |
| | 대학원 재학 또는 졸업 | 56 | 11.2 |
| 주관적 사회경제적지위 | 최하 | 11 | 2.2 |
| | 하 | 43 | 8.6 |
| | 중하 | 138 | 27.6 |
| | 중 | 205 | 41 |
| | 중상 | 86 | 17.2 |
| | 상 | 17 | 3.4 |
| | 최상 | 0 | 0 |
| Total | | 500 | 100 |

적 요인 분석을 위한 구조방정식 모형(structural equation modeling) 검증을 위해서는 최소 200명 이상의 참가자를 모집하는 것이 요구되며(Tomarken & Waller, 2005), 개발된 문항 10배수 이상의 참가자를 모집할 것을 권장한다(Hair et al., 2010). 따라서 개발한 척도의 타당화를 위해 생성 인공지능 서비스를 이용한 사람들을 대상으로 했으며 측정도구의 일반화를 위해 세대별 100명의 참가자를 모집해 총 500명의 참가자를 모집했다. 확인적 요인 분석은 탐색적 요인 분석과 동일하게 R 프로그램을 사용하였으며, 분석 패키지는 lavaan 패키지를 사용했다(Rosseel, 2012). 모형 요인 구조 적합도는 Tucker-Lewis

Index(TLI; Bentler & Bonett, 1980), Competitive Fit Index(CFI; Bentler, 1990), SRMR(Hu & Bentler, 1999), 그리고 RMSEA(Steiger, 1990)를 통해 검증했다.

3.4.2. 연구 결과

확인적 요인 분석 결과 네 요인 모형의 모형 적합도가 분석을 수행하기에 통계적으로 적합하다는 것을 확인할 수 있었다. 구체적으로 CFI와 TLI 값의 경우 각각 .960과 .945로 기준치(.90 이상)에 부합하는 것을 확인할 수 있었다. SRMR 값은 .042로 기준치(.80 이하)에 부합하였으며, RMSEA 값도 .078로 기준치(.80 이하)에 부합한다는

〈표 4〉 확인적 요인 분석 모형 적합도 분석 결과

| Fit Index | CFA | Recommendation |
|----------------|-------------------|----------------|
| CFI | .960 | >.90 |
| TLI | .945 | >.90 |
| SRMR | .042 | <.08 |
| RMSEA [90% CI] | .078 [.067, .090] | <.08 |

〈표 5〉 확인적 요인 분석을 통해 측정도구 요인 부하량 및 적합성 검증

| | <i>b</i> | Std.Err | z-value | P(> z) | β |
|------------|----------|---------|---------|---------|---------|
| AI 활용능력 <- | | | | | |
| A1 | 1 | | | | 0.852 |
| A2 | 1.097 | 0.043 | 25.43 | <.001 | 0.897 |
| A3 | 1.119 | 0.051 | 21.998 | <.001 | 0.827 |
| 비판적 평가 <- | | | | | |
| B1 | 1 | | | | 0.773 |
| B2 | 1.065 | 0.054 | 19.801 | <.001 | 0.855 |
| B3 | 0.974 | 0.065 | 14.974 | <.001 | 0.731 |
| 윤리적 사용 <- | | | | | |
| C1 | 1 | | | | 0.796 |
| C2 | 1.026 | 0.055 | 18.622 | <.001 | 0.856 |
| C3 | 0.623 | 0.057 | 10.863 | <.001 | 0.500 |
| 창의적 활용 <- | | | | | |
| D1 | 1 | | | | 0.797 |
| D2 | 1.098 | 0.054 | 20.41 | <.001 | 0.863 |
| D3 | 1.055 | 0.052 | 20.256 | <.001 | 0.845 |

것을 확인할 수 있었다(Kline, 2015). 이와 같은 분석 결과는 생성 인공지능 리터러시 척도의 네 요인 모형 구조가 통계적으로 적합하다는 것을 의미한다.

또한 네 가지 잠재 변인과 측정도구 간의 관계성 확인을 위해 요인 부하량을 확인한 결과 각 잠재 변인의 측정 도구는 잠재 변인을 측정하기에 적합한 도구임을 확인할 수 있었다. 표준화된 부하량은 한 측정변인을 제외하고 모두 .70 이상임을 확인할 수 있었다. 윤리적 사용의 세 번째 도구의 경우 요인 부하량 표준화 값이 .50으로 나타나 설명력이 약한 변인으로 보일 수 있으나, 이 문항을 제거한 모형의 경우 오히려 모형 적합도에 부정적인 영향을 미쳤다(SRMR=.044; RMSEA=.084). 다만, 세 번째

문항의 설명력이 적고($R^2=.250$) 신뢰도 수치가 수용할 수 있는 수준이지만 충분하지는 않은 것으로 나타났기에 이 문항에 관한 후속 연구가 필요할 것이다. 20-30대를 대상으로 한 분석에서도 모형 적합도 수준에서 약간의 개선 효과가 나타났기에 연령에 따른 효과를 함께 고려한 후속 연구가 필요할 것으로 판단된다. 또한 잠재 변인의 분산 중 음(-)의 값을 갖는 경우가 발견되지 않았다. 따라서 측정도구를 구성하는 차원에서 본 연구에 심각한 문제는 없는 것으로 판단했다.

끝으로 측정도구의 신뢰도를 확인하기 위하여 신뢰도 분석을 수행했다. 신뢰도 분석은 세 가지 기준을 통해 확인했다. 우선 기본적인 내적 신뢰도 수준을 Cronbach's

〈표 6〉 생성 인공지능 리터러시 네 가지 하위 요인 신뢰도 분석 결과

| | AI 활용능력 | 비판적 평가 | 윤리적 활용 | 창의적 활용 |
|-------|---------|--------|--------|--------|
| Alpha | .890 | .818 | .754 | .872 |
| CR | .893 | .828 | .767 | .874 |
| AVE | .735 | .616 | .534 | .699 |

〈표 7〉 생성 인공지능 리터러시 하위 요인 간 상관 분석 결과 및 AVE 제곱근 값 비교

| | 1 | 2 | 3 | 4 |
|------------|-------------|-------------|-------------|-------------|
| 1. AI 활용능력 | .857 | | | |
| 2. 비판적 평가 | .629 | .785 | | |
| 3. 윤리적 활용 | .548 | .805 | .731 | |
| 4. 창의적 활용 | .689 | .622 | .692 | .836 |

Note. Diagonal values in bold represent the square root of the AVE (Average Variance Extracted).

〈표 8〉 HTMT를 이용한 변별 타당성 분석 결과

| | 1 | 2 | 3 | 4 |
|------------|------|------|------|---|
| 1. AI 활용능력 | — | | | |
| 2. 비판적 평가 | .665 | — | | |
| 3. 윤리적 활용 | .552 | .799 | — | |
| 4. 창의적 활용 | .705 | .658 | .704 | — |

alpha 값(.70 이상)으로 확인했고, 복합 신뢰도(composite reliability, CR; .70 이상)와 평균 추출 분산(average variance extracted, AVE; .50 이상) 값을 확인했다(Bagozzi & Yi, 1988). 분석 결과 네 가지 잠재 변인 모두 alpha 값이 .70 이상이었으며, 복합 신뢰도와 평균 추출 분산 또한 기준을 충족한다는 것을 확인할 수 있었다. 따라서 확인적 요인 분석과 신뢰도 분석을 통해 생성 인공지능 리터러시 척도의 타당도와 신뢰도를 검증할 수 있었다. 네 요인 간 상관 계수를 확인한 결과에서도 모든 변인이 정적 관계성을 가진다는 것을 확인할 수 있었다. 이와 같은 연구 결과는 생성 인공지능 리터러시를 측정하기 위해 개발된 네 가지 요인의 모형 구조가 통계적으로 적합하다는 것을 의미한다.

본 연구에서는 Fornell-Larcker 기준을 사용하여 변별 타당성을 검증하였다. 각 차원의 AVE의 제곱근을 다른

차원 간 상관계수와 비교한 결과, 대부분의 차원에서 변별타당성이 확보되었음을 확인하였다. 그러나 ‘비판적 평가’와 ‘윤리적 활용’ 차원 간 상관계수는 .805로 나타나, 두 차원 간 변별타당성이 충분히 확보되지 못한 것으로 보인다. 이는 두 차원이 모두 인공지능 사용에 대한 윤리적 책임과 관련된 특성을 다루고 있어, 개념적 유사성으로 인해 상관성이 높아진 것으로 해석된다. 따라서 추가적인 검증을 위해 Heterotrait-monotrait ratio (HTMT) 분석을 수행했다. 분석 결과 .850 이상의 값을 갖지 않는 것으로 나타나 판별 타당성이 충족된다는 것을 확인할 수 있었다(Henseler et al., 2015).

3.4.3. 측정도구 타당성 분석 결과

본 연구에서 개발한 네 요인으로 구성된 인공지능 리터러시 척도의 타당도를 확인하기 위해 각 하위 요인 별

평균값을 구한 변인과 타당성 분석을 위해 측정된 변인 간의 상관 분석을 수행했다. 상관 분석은 기존 연구에서 개발한 인공지능 리터러시 척도(Wang et al., 2023), 그리고 인공지능을 비롯한 기술에 대한 인식 수준을 확인하기 위해 기술 두려움 인식(Khasawneh, 2018), 인공지능 서비스 친숙성(Chi et al., 2021), 인공지능 서비스 이용 효능감(White et al., 2012), 그리고 인공지능 서비스에 대한 지각된 위험(Venkatesh et al., 2012)을 측정하고 분석에 사용했다. 이 도구는 확인적 요인 분석에 이어서 타당화 분석을 위해 확인적 요인 분석 데이터 수집 과정에서 함께 측정되었다.

우선 기존 인공지능 리터러시 척도와 개발한 척도 간의 상관 분석을 수행한 결과 각 하위 요인이 기존 척도와 .3 이상에서 .7 미만의 상관성을 갖는다는 것을 확인할 수 있었다. 이는 기존 척도와 현 개발된 척도가 관계성이 매우 낮거나 매우 유사한 수준의 문항이 아닌 서로 구분되는 척도임을 의미한다. 이를 통해 변별 타당성이 있다는 것을 확인할 수 있었다. 기존 척도와의 상관성 중 인공지능에 대한 인지 수준은 비판적 평가와 상관성이 가장 높았고($r=.595, p<.001$), 인공지능 사용은 AI 활용능력과 상관성이 높게 나타났다($r=.685, p<.001$). 또한 인공지능 평가의 경우 창의적 활용과 상관성이 가장 높았으며($r=.663, p<.001$), 윤리적 사용은 인공지능 윤리($r=.564, p<.001$) 요인과 상관성이 높게 나타났다. 이를 통해 기존 척도와 현 척도가 각 하위 요인 간 관계성을 갖지만 서로

구분되는 개념임을 확인할 수 있었다.

다음으로 기술 두려움에 대한 상관 분석을 수행한 결과 AI 활용능력은 유의한 상관관계가 나타나지 않았다($r=.019, p=.668$). 다른 요인은 약 .20 정도의 유의한 상관성을 갖는다는 것을 확인할 수 있었다. AI 서비스 친숙성의 경우 AI 활용능력은 상관관계가 유의하지 않았으며, 비판적 평가와 창의적 활용은 .50 이상의 유의한 상관성을 갖는다는 것을 확인할 수 있었다. 반면 지각된 위험의 경우 윤리적 사용이 가장 높은 상관관계가 나타났고($r=.191, p<.001$), 비판적 평가($r=.164, p<.001$)와 창의적 활용($r=.179, p<.001$)도 상관관계가 통계적으로 유의했다. 다만 AI 활용능력은 지각된 위험과도 상관관계가 통계적으로 유의하지 않았다. AI 이용 효능감과의 상관 분석 결과 AI 서비스 친숙성과 유사한 수준으로 상관관계가 확인되었다. 개발한 생성 인공지능 리터러시 측정도구와 다른 도구 간 전반적인 상관관계는 <표 9>를 통해 확인할 수 있다. 이를 통해 기존에 생성 인공지능 혹은 인공지능 기술에 대한 사람들의 태도와 현재 개발된 인공지능 리터러시 척도 중 관련된 하위 요인이 유의한 관계성을 갖는다는 것을 확인할 수 있었다.

끝으로 인공지능 서비스에 대한 지각된 효율성(Kang et al., 2023)에 대해 생성 인공지능 리터러시의 네 가지 하위 요인이 미치는 상대적 영향력 수준을 확인하기 위해 회귀분석을 수행했다. 분석 결과 생성 인공지능 리터러시를 구성하는 네 요인은 지각된 효율성을 36.3% 설명

<표 9> 측정도구 타당성 확인을 위한 상관 분석 결과

| | AI 인식 | AI 사용 | AI 평가 | AI 윤리 | 기술 두려움 | AI 친숙성 | AI 효능감 | 지각된 위험 |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| AI 활용능력 | .587*** | .685*** | .638*** | .337*** | .019 | .509*** | .531*** | .028 |
| 비판적 평가 | .595*** | .611*** | .655*** | .483*** | .167*** | .514*** | .526*** | .164*** |
| 윤리적 사용 | .499*** | .498*** | .553*** | .564*** | .202*** | .458*** | .430*** | .191*** |
| 창의적 활용 | .551*** | .653*** | .663*** | .406*** | .197*** | .584*** | .584*** | .179*** |

*** $p<.001$

했다. 또한 AI 활용능력($b=.295, p<.001$)과 창의적 활용($b=.223, p<.001$)이 미치는 영향은 통계적으로 유의했으며, 다른 두 변인은 그렇지 않았다($bs=.06, n.s.$).

인공지능에 대한 불안 인식(Wang & Wang, 2022)에 대해서도 동일한 분석을 수행했다. 생성 인공지능 리터러시는 인공지능에 대한 불안 인식을 5.8% 설명했다. 이는 분석 결과에서도 나타났는데 윤리적 사용이 불안 인식에 미치는 영향이 통계적으로 유의했고($b=.216, p=.003$), 비판적 평가가 미치는 영향은 한계적으로 유의했다($b=.138, p=.073$). 다른 두 변인은 통계적으로 유의하지 않았다. 불안에 대한 인식은 전반적으로 설명력이 적었으나 이는 윤리나 비판적 평가와 같은 부분 이외에 사용 역량이나 창의적 활용이 이와 같은 분석을 통해 생성 인공지능에 대한 사람들의 인식이나 평가를 예측하는 생성 인공지능 리터러시 척도와 각 하위 요인이 미치는 영향에 대한 분석을 통해 척도의 예측력과 타당성을 확인할 수 있었다. 이를 통해 본 연구에서 개발한 척도가 생성 인공지능에 대한 소비자의 평가와 인식 수준을 예측한다는 것을 확인할 수 있었다. 이와 같은 과정을 통해 본 연구에서 개발한 척도의 타당성을 확인했다.

4. 종합 토의

4.1. 연구 결과 요약

생성 인공지능의 등장 이후 인공지능 기술은 더 이상 미래가 아닌 현실이 되었다. 그에 따라 새로운 기술인 생성 인공지능에 대해서도 사람들이 이를 이해하고 활용할 수 있는 역량을 키워야 한다는 목소리가 높아진다(Wang et al., 2023). 일반적으로 새로운 기술은 사용하는 방법이나 역량을 익혀 원하는 결과를 얻을 수 있는 능력을 키우는 것이 중요하며 이와 같은 개념을 디지털 리터러시의

차원으로 정의된다(Aydin, 2021). 그러나 생성 인공지능은 다른 기술과 달리 자연어를 기반으로 쉬운 명령어를 사용해 원하는 결과를 얻을 수 있다는 측면에서 기술을 사용하는 것에 대해 새로운 역량이나 능력을 키우지 않아도 된다는 차이가 있다(Orru et al., 2023). 다만 쉬운 사용으로 인해 일부 이용자가 생성 인공지능을 자신이 원하는 방향으로 악용하거나, 부정적인 결과물을 일부러 산출하기 위한 피해 발생 사례가 지속적으로 나타나 주의를 요구한다(Ghallab, 2019; Lund et al., 2023). 따라서 생성 인공지능을 활용하기 위해서는 기존의 사용 역량을 높이는 데 초점을 맞추는 것이 아니라 생성 인공지능을 사용하는 과정에서 발생할 수 있는 윤리적인 부분을 고려하는 것이 필요하다. 이에 본 연구에서는 기존 디지털 리터러시와 인공지능 리터러시를 측정하기 위해 도구를 제안한 선행 연구(List, 2019; Wang et al., 2023)를 바탕으로 생성 인공지능 리터러시를 측정하기 위한 도구를 제안하고자 했다. 기존의 연구가 주로 새로운 기술에 대한 사용 역량과 기술에 대한 이해에 중점을 둔 반면 생성 인공지능의 자연어 기반 처리가 갖는 특징에 따라 윤리적인 부분이 강조된다(Orru et al., 2023). 따라서 본 연구에서도 윤리적인 사용에 대한 고려와 함께 효과적인 인공지능 사용을 위한 창의적 활용과 생산된 정보에 대한 비판적 평가를 함께 고려한 연구를 수행하고자 했다. 이를 위해 생성 인공지능 리터러시 측정 도구의 하위 요인은 인공지능에 대한 이해와 사용 능력을 다룬 'AI 활용능력', 새로운 개념이나 정보를 생산하고 문제를 해결하는 '창의적 활용', 정보에 대한 평가와 비판적 시각을 다룬 '비판적 평가' 그리고 인공지능 사용 과정에서 발생하는 윤리적 문제를 고려한 '윤리적 사용'의 네 가지 요인으로 구성했다. 탐색적 요인 분석과 확인적 요인 분석을 수행한 결과 처음 개발된 32개의 문항 중 최종적으로 12개의 문항, 네 가지 요인을 구성하는 측정도구가 타당하다는 것을 확인할 수 있었다. 측정 도구의 신뢰도도 일관적으

로 나타났으며, 기존 척도 및 인공지능 관련 평가 척도와 의 상관 분석에서도 의미 있는 관계성을 확인할 수 있었다. 이 결과는 기존의 활용 능력과 정보에 대한 평가가 강조되었던 인공지능 리터러시가 생성 인공지능을 사용하는 과정에서 중요한 윤리적 부분을 추가로 고려했다는 점에서 의의가 있다고 할 수 있다.

4.2. 이론적 및 실무적 함의

본 연구는 기존의 정보를 읽고 처리하는 역량을 의미하는 리터러시의 개념을 디지털 리터러시로 확장한 현대 사회에서 인공지능 시대에 부합하는 생성 인공지능 리터러시를 측정하고자 했다. 이 과정에서 탐색적 요인 분석에서는 생성 인공지능을 들어봤거나 사용해 본 적이 있는 사람들을 대상으로 측정 도구를 구성하기 위한 분석을 수행했다. 이는 기존의 인공지능 리터러시 관련 측정 도구가 특정 세대가 계층을 위한 도구를 개발하거나(Ng et al., 2024), 문제 해결 및 창의적 활용에 대한 부분이 간과되거나(Wang et al., 2023), 이용과 활용 측면이 강조되었다(Carolus et al., 2023). 이를 본 연구에서는 대한민국 성인을 대상으로 도구 개발 연구를 수행해 전 연령을 포괄하는 일반화된 측정도구를 만들고자 했다. 이와 같은 과정을 통해 생성 인공지능 리터러시 영역을 개인에 따른 차이를 확인하는 도구로 타당화하기 위한 확인적 요인 분석을 수행했다. 이때 실제 인공지능 서비스를 이용해 본 적이 있는 사람들을 대상으로 생성 인공지능 리터러시 척도가 실제 사람들의 인공지능 서비스에 대한 평가를 예측할 수 있는지를 확인하고 기존 척도와의 관계성을 검증하고자 했다. 이를 통해 본 연구는 생성 인공지능 리터러시에 대한 구성 개념을 제안하고, 전 연령을 대상으로 한 일반화 척도를 개발하고자 했으며, 기존의 연구가 주로 서양권의 사람들을 대상으로 연구가 이루어져 한국의 이용자들에게 적합한 도구를 만들 필요를 충

족했다는 차원에서 의미가 있다.

생성 인공지능은 자연어를 통해 활용 방법을 쉽고 빠르게 학습할 수 있지만, 사람들이 갖는 편견이나 고정관념으로 인해 쉽게 이용하지 못하고 있다는 것을 최근 패널조사 보고를 통해 확인할 수 있다(방송통신위원회, 2024). 조사 결과에 따르면 생성 인공지능을 이용해 본 적이 있는 사람은 전체 패널 중 12.3%에 그쳤으며 주된 활용도 텍스트 생성에 머물렀다. 또한 프라이버시 보호를 위한 행동을 생성 인공지능을 사용하는 과정에서 얼마나 수행했는가에 대한 비율에서도 대부분 50%를 넘지 못하는 등 윤리적 차원과 창의적 차원의 활용이 잘 이루어지지 못하고 있다는 것을 확인할 수 있다. 이 결과는 생성 인공지능 관련 서비스를 이해하고 활용하는 능력을 넘어서야 함을 보여준다. 특히, 이를 창의적으로 활용하고, 생성된 정보를 비판적으로 평가하며, 과정에서 발생할 수 있는 윤리적 문제를 적절히 다루는 것이 중요하다. 따라서 본 연구에서 개발한 도구를 바탕으로 정기적인 패널조사와 소비자 만족도 조사를 수행하며 장기적 차원에서 얻은 데이터를 통해 실제 시민들의 리터러시 수준을 확인하고 부족한 부분을 보강하기 위한 교육 및 정책 프로그램을 개발하는 것이 필요할 것이다.

더불어 기업이 서비스를 제공하는 차원에서도 이용자들의 만족도와 활용성을 높이기 위해 창의적 활용과 윤리적 사용 그리고 생산한 정보를 악용하지 않고 한 번 더 검토할 수 있는 비판적 평가 요인 수준을 측정하고 관리하는 차원에서도 이 도구를 사용할 수 있을 것이다(이경주, 김은영, 2020). 전반적인 리터러시 수준이 높을수록 생성 인공지능을 통해 더욱 많은 자원을 확보하고 그 활용성을 넓힐 수 있을 것이므로 서비스 이용 과정에서 경험하는 수준에서도 차이가 발생할 것이다. 따라서 기존 이용자들의 리터러시 수준을 측정하고 그들에게 맞는 서비스를 제공함으로써 이용자들의 리터러시 향상과 서비스 지속 사용 의도를 높이는 방향으로 소비자 교육

및 마케팅적 활용도 가능할 것이다. 또한 생성 인공지능이 등장한 초기에 가장 주의 깊게 다루어진 문제는 생성 인공지능을 통해 생산된 정보가 실제 존재하지 않는 가상의 정보나 실제 사실에 기반하지만 정확하지 않은 정보를 생산해 이용자에게 제공하는 환각(hallucinations) 정보 현상이 나타났다(Chelli et al., 2024). 이는 생성 인공지능이 기본에 학습한 정보를 바탕으로 새로운 정보를 생산할 때 토근 단위로 정보를 생산함에 따라 새로 문장을 만드는 과정에서 정확하지 않은 정보가 섞여 생산되는 문제가 되었다. 이는 서비스 자체가 큰 문제가 있다기 보다는 생성 인공지능의 특징이 묻어나는 것으로 생성 인공지능을 통해 생산된 정보를 다시 비판적인 시각에서 판단하고 사실관계를 확인하기 위한 노력이 필요하다고 볼 수 있다. 왜냐하면 이 특징을 없애는 것은 기존의 정보 검색 서비스와 큰 차이가 없으며, 생성 인공지능을 통해 원하는 정보를 명확하고 빠르게 정리할 수 있기 때문이다. 따라서 본 연구에서 개발한 비판적 평가 요인을 바탕으로 이용자의 비판적 평가 수준을 확인하고 서비스 이용 과정에서 자칫 잘못된 정보를 전달하는 이용자가 되지 않을 수 있도록 리터러시 수준에 따른 교육이나 가이드라인을 제시할 수 있을 것이다.

4.3. 연구 한계 및 향후 연구 방향

본 연구는 디지털 리터러시 척도를 인공지능 리터러시 척도로 확장하고 새로운 개념을 제안한 선행 연구를 바탕으로 한국형 생성 인공지능 리터러시 척도를 개발하고 이를 타당화 하기 위한 목적에서 진행되었다. 이 과정에서 생성 인공지능 리터러시를 이용할 때 고려해야 할 윤리적 사용에 대한 부분과 함께 창의적 활용, 비판적 평가의 요인을 제안해 생성 리터러시 개념을 확장하고자 했다. 그러나 다음과 같은 차원에서 보완이 필요하다. 우선 윤리적 차원의 요인이 현재의 측정 도구 분석 결과에서

설명력이 약하다는 단점이 있다. 생성 인공지능을 사용하는 다양한 환경을 고려해 윤리적인 부분을 얼마나 고려하는가를 측정하고자 한 본 연구의 도구는 생성 인공지능에 대한 개인의 지식이나 인식 수준, 사용해 본 서비스나 그 서비스를 통해 얻은 결과, 윤리적인 부분에 대한 고려나 개인의 윤리적 민감성 수준 등을 고려하지 못해 어떤 사람들에게서 차이가 발생한 것인지를 명확히 확인하지 못했다. 또한 설명력이 낮은 문항의 경우 그 문항 자체의 문제인지, 아니면 아직 그 문항이 의미하는 바와 같은 다양한 사회적 문제나 논의가 이루어지지 못한 것인지에 대해서도 현시점에서 명확한 결론을 내리기 어렵다. 따라서 후속 연구에서는 가상의 시나리오 상황이나 특정 문제에 대한 사람들의 윤리적, 도덕적 평가 상황을 가정해 개인의 리터러시 수준에 따른 도덕적, 윤리적 평가 차이를 확인하는 연구를 수행할 수 있을 것이다. 이 과정에서 측정 도구의 타당성을 평가할 뿐만 아니라 참여자의 인구통계를 명확히 구분하고, 생성 인공지능에 대한 경험이나 학습 수준을 함께 고려해 개인에 따른 차이가 발생하는 요인을 명확히 구분하는 연구를 수행할 수 있을 것이다.

다음으로 생성 인공지능의 발전이 매우 빠르게 이루어지고 있는 상황에서 현시점에서 개발된 도구가 앞으로 몇 년 아니 몇 달 안으로 소개될 새로운 생성 인공지능 서비스나 생성 인공지능을 활용한 다양한 ICT 관련 서비스에 대해 명확한 평가나 개인의 리터러시 수준을 측정할 수 있는 도구인지가 확실하지 않다. 따라서 새로운 서비스나 생성 인공지능에 대한 새로운 관점이나 요인이 도출되면 이를 보완한 후속 도구 개발이 함께 이루어져야 할 것이다. 다만, 현시점에서 생성 인공지능을 단순히 개인의 역량으로만 보는 것이 아니라 이 도구를 통해 더욱더 효과적이고 효율적이며 윤리적인 차원까지 고려한 도구를 개발하고자 했으므로 이와 같은 요인을 고려하면 새로운 서비스나 윤리적 차원의 문제에 대한 개인

의 수준 차이를 측정하는 도구로 활용할 수 있을 것이다.

본 연구에서는 한국형 생성 인공지능 리터러시 척도를 개발하였으나, 문화적 특성, 이용 맥락, 서비스의 특징과 같은 한국적 요소를 충분히 반영하는 데에는 한계가 있었다. 현재 한국에서의 생성 인공지능 관련 서비스와 이용 환경에 대한 연구가 부족한 상황이므로, 이러한 요소들이 척도에 충분히 반영되지 않았을 가능성이 있다. 특히, 생성 인공지능 기술이 한국에서 어떻게 수용되고 활용되는지, 그리고 이에 따른 문화적, 사회적 맥락이 어떠한지를 더욱 깊이 있게 분석하는 추가 연구가 필요하다. 이러한 연구를 통해 한국형 척도가 더욱 정교하게 발전될 수 있을 것이며, 한국적 특성을 반영한 척도를 통해 연구의 신뢰성과 타당성을 더욱 높일 수 있을 것이다. 앞으로의 연구에서는 한국적 이용 맥락과 문화적 요소를 보다 심도 있게 다루어, 생성 인공지능 리터러시 척도가 한국인의 실질적인 경험과 인식을 더욱 정확하게 반영할 수 있도록 하는 방향으로 나아가야 한다. 우선, 본 연구에서 제시된 4개의 리터러시 차원(‘AI 활용능력’, ‘비판적 평가’, ‘윤리적 사용’, ‘창의적 활용’)은 한국적 맥락을 반영하여 개발되었으나, 이러한 차원이 실제로 한국 사회에서 다른 문화권과 비교했을 때 얼마나 독특한지에 대한 심층적인 비교 연구는 부족하다. 한국 사회의 디지털 인프라와 AI 기술에 대한 높은 적응력을 고려할 때, 이러한 차원들이 다른 국가에서도 유사한 방식으로 적용될 수 있는지를 확인하기 위한 추가 연구가 필요하다. 특히, ‘창의적 활용’과 ‘윤리적 사용’ 차원은 한국에서 강조된 개념이지만, 다른 국가에서도 동일하게 중요한 역할을 하는지를 분석할 필요가 있다.

또한, 본 연구는 표본을 기반으로 한국 내 사용자들을 대상으로 척도를 개발하고 타당성을 검증했으나, 다양한 인구 통계적 특성(연령, 성별, 교육 수준 등)에 따른 AI 리터러시 차원의 차이를 더욱 명확하게 분석하지 못했다. 향후 연구에서는 세부 인구 집단 간의 차이를 분석하

여, 특정 집단에서 특히 강화되어야 할 리터러시 영역을 파악하는 것이 필요하다. 예를 들어, 젊은 세대는 ‘창의적 활용’에 더 큰 잠재력을 보일 수 있지만, 고령층은 ‘윤리적 사용’에 대한 교육이 더 필요할 수 있다.

마지막으로, 본 연구는 AI 리터러시 척도의 개발과 타당성을 확인하는 데 중점을 두었으나, 실제 교육 현장이나 AI 관련 서비스에서 이 척도를 적용해 그 효과를 검증하는 과정이 아직 진행되지 않았다. 따라서 향후 연구에서는 이 척도를 교육 프로그램이나 기업의 AI 윤리 교육에서 활용하여 리터러시 수준을 실제로 향상할 수 있는지, 그리고 이러한 척도가 실제 문제해결이나 윤리적 판단에 어떻게 이바지하는지 평가하는 실증 연구가 필요하다.

통계적인 관점에서도 다음과 같은 한계를 갖는다. 첫째, 탐색적 요인분석(EFA) 과정에서 네 가지 요인(AI 활용능력, 창의적 활용, 비판적 평가, 윤리적 사용)을 바탕으로 한 요인 구조를 설정하였으나, 이론적 구조와 선행 연구에 기반한 선택임에도 불구하고, 실제 분석 결과에서 일부 요인 간의 상관관계가 높게 나타났다. 특히, ‘비판적 평가’와 ‘윤리적 활용’ 요인 간의 상관관계가 .805로 높게 나타나, 두 요인 간의 변별 타당성 확보에 한계가 있었다. 그러나 HTMT 분석 결과 .850 이하로 나타나 변별 타당성이 검증된 것으로 확인되었다. 다만 상관관계가 높다는 것은 추가적인 연구가 필요하다는 것을 의미한다. 따라서 향후 연구에서는 두 차원의 개념적 구분을 명확히 하기 위해 문항 구성을 보다 정교화하거나, 추가적인 검증 도구를 통해 차원을 재정립할 필요가 있다.

둘째, 설문 설계 및 자료 수집 과정에서 CMV(Common Methods Bias)와 다중공선성(multicollinearity) 문제를 최소화하기 위해 다양한 전략을 사용하였으나, 이러한 편향을 완전히 배제할 수 없다는 점도 연구의 한계로 작용할 수 있다. 각 요인의 하위 문항들은 다른 요인들과 혼합하여 배치되었고, 중복되지 않도록 설계되었지만, 응

답자의 인식이나 답변 방식에 따라 편향이 발생할 가능성은 여전히 존재한다. 이러한 한계점을 고려하여 후속 연구에서는 더욱 정교한 설계와 분석을 통해 이러한 문제를 보완할 필요가 있다. 이와 같은 한계점을 바탕으로, 본 연구에서 개발된 척도가 한국적 맥락에서 AI 리터러시를 평가하는 데 중요한 도구로 기능할 수 있음을 인정하면서도, 후속 연구에서 더욱 정밀한 분석과 검토를 통해 연구 타당성과 신뢰성을 높여야 할 것이다.

끝으로, 본 연구에서는 생성 인공지능 리터러시를 평가하기 위해 다양한 변수들을 사용하여 상관관계 분석을 수행했다. 이 과정에서 각 변수 간의 관계를 탐색함으로써 연구의 타당성을 높이고자 하였으나, 상관관계 분석이 연구의 주요 결론을 도출하는 데 있어 절대적인 역할을 하기에는 여전히 한계가 있다. 상관관계 분석은 변수 간의 관계를 파악하는 데 유용하지만, 이러한 관계가 반드시 인과관계를 나타내지는 않는다. 또한, 사용된 변수들에 대한 적합성 검증은 연구 초기 단계에서 선행연구를 기반으로 이루어졌으나, 상관관계 분석을 통해 측정 도구의 타당성을 검증하기 위해서는 추가적인 검토와 분석이 필요하다. 후속 연구에서는 이러한 한계점을 개선하기 위해 모형 검증 방식을 적용하여 연구의 신뢰성과 타당성을 강화할 필요가 있을 것이다.

〈참고문헌〉

[국내 문헌]

1. 김세나, 표두미, 이지선, 이준기, 신강현, 김경일 (2015). 한국판 조절초점 척도 타당화 연구. *한국심리학회지: 사회 및 성격*, *29*(3), 85-110.
2. 류미영, 한선관 (2023). 인공지능 리터러시 측정을 위한 검사 지표 개발. *한국컴퓨터정보학회논문지*, *28*(7), 39-46.
3. 이경주, 김은영 (2020). 플랫폼 서비스 혁신에 있어 인공지능(AI)의 역할과 효과에 관한 연구: 카카오 그룹의 인공지능 활용 사례 연구. *지식경영연구*, *21*(1), 175-195.
4. 이병관, 윤태웅, 노환호, 임혜빈 (2019). 한국형 사회적 배제 경험 척도 개발 및 타당화 연구. *한국심리학회지: 소비자·광고*, *20*(1), 127-152.
5. 자유첸, 백지연, 조현정, 이지연 (2022). 사무직 직원의 디지털 리터러시와 무형식 학습의 관계에서 인센티브의 조절효과에 대한 연구: 중국의 사례를 중심으로. *지식경영연구*, *23*(3), 173-192.
6. 최숙영 (2022). AI 리터러시 프레임워크에 대한 연구. *컴퓨터교육학회 논문지*, *25*(5), 73-84.
7. 탁진국 (2007). *심리검사 -개발과 평가방법의 이해-* (2판). 학지사.
8. 황용석, 이선민, 김여림, 황현정 (2022). 디지털 역량 척도 개발 연구. *언론정보연구*, *59*(2), 5-48.
9. Acharya, N., Sassenberg, A. M., & Soar, J. (2023). Effects of cognitive absorption on continuous use intention of AI-driven recommender systems in e-commerce. *Foresight*, *25*(2), 194-208.
10. Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211-36.
11. Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management*, *62*, 102433.
12. Aydin, M. (2021). Does the digital divide matter? Factors and conditions that promote ICT literacy. *Telematics and Informatics*, *58*, 101536.
13. Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608-1613.
14. Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74-94.
15. Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246.
16. Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588-606.
17. Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, *71*, 499-515.
18. Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*, *8*(1), 108-117.
19. Brown, J. A. (1998). Media literacy perspectives. *Journal of Communication*, *48*(1), 44-57.
20. Buckingham, D., & Burn, A. (2007). Game literacy in theory and practice. *Journal of Educational Multimedia and Hypermedia*, *16*(3), 323-349.
21. Butler, J. (2012). Grappling with change: Web 2.0 and teacher education. In D. Polly, C. Mims, & K. A. Perschitte (Eds.), *Developing technology-rich teacher education programs: Key issues* (pp. 135-150). Hershey, PA: IGI Global.
22. Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends*, *68*(1), 47-57.
23. Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS—Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change—and meta—competencies. *Computers in Human Behavior: Artificial Humans*, *1*(2), 100014.
24. Celik, I. (2023). Exploring the determinants of artificial intelligence (Ai) literacy: Digital divide, computational

- thinking, cognitive absorption, *Telematics and Informatics*, **83**, 102026.
25. Celik, I., Dindar, M., Muukkonen, H., & Jarvela, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research, *TechTrends*, **66**(4), 616–630.
 26. Chai, C. S., Wang, X., & Xu, C. (2020). An extended theory of planned behavior for the modelling of Chinese secondary school students' intention to learn artificial intelligence, *Mathematics*, **8**(11), 2089.
 27. Chelli, M., Descamps, J., Lavoue, V., Trojani, C., Azar, M., Deckert, M., ... & Ruetsch-Chelli, C. (2024). Hallucination rates and reference accuracy of ChatGPT and bard for systematic reviews: Comparative analysis, *Journal of Medical Internet Research*, **26**, e53164.
 28. Chen, D., Wu, J., & Wang, Y. (2011). Unpacking new media literacy, *Journal on Systemics, Cybernetics and Informatics*, **9**, 84–88.
 29. Chi, O. H., Jia, S., Li, Y., & Gursoy, D. (2021). Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery, *Computers in Human Behavior*, **118**, 106700.
 30. Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world, *Harvard Business Review*, **96**(1), 108–116.
 31. DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Thousand Oaks, CA: Sage.
 32. Duffy, B. (2018). *The perils of perception*. London, UK: Atlantic Books.
 33. Dunn, A. G., Shih, I., Ayre, J., & Spallek, H. (2023). What generative AI means for trust in health communications, *Journal of Communication in Healthcare*, **16**(4), 385–388.
 34. Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration, *Journal of Information Technology Case and Application Research*, **25**(3), 277–304.
 35. Ghallab, M. (2019). Responsible AI: Requirements and challenges, *AI Perspectives*, **1**(1), 1–7.
 36. Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines, *Minds and Machines*, **30**(1), 99–120.
 37. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Englewood Cliffs: Prentice Hall.
 38. Hetzel, R. D. (1996). A primer on factor analysis with comments of practices and reporting. In B. Thompson (Ed.), *Advances in Social Science Methodology*, Vol. 4, (pp.175–206). Greenwich, CT: JAL.
 39. Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling, *Journal of the Academy of Marketing Science*, **43**, 115–135.
 40. Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling: A Multidisciplinary Journal*, **6**(1), 1–55.
 41. Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions, *Computers in Human Behavior*, **80**, 295–302.
 42. Jiang, J., Karran, A. J., Coursaris, C. K., Leger, P. M., & Beringer, J. (2023). A situation awareness perspective on human-AI interaction: Tensions and opportunities, *International Journal of Human-Computer Interaction*, **39**(9), 1789–1806.
 43. Kang, E. Y. N., Chen, D. R., & Chen, Y. Y. (2023). Associations between literacy and attitudes toward artificial intelligence-assisted medical consultations: The mediating role of perceived distrust and efficiency of artificial intelligence, *Computers in Human Behavior*, **139**, 107529.
 44. Khasawneh, O. Y. (2018). Technophobia: Examining its hidden factors and defining it, *Technology in Society*, **54**, 93–100.
 45. Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.
 46. Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies, *Computers and Education: Artificial Intelligence*, **6**, 100225.

47. Koc, M., & Barut, E. (2016). Development and validation of New Media Literacy Scale (NMLS) for university students. *Computers in Human Behavior*, *63*, 834–843.
48. Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2022). Evaluating artificial intelligence literacy courses for fostering conceptual learning, literacy and empowerment in university students: Refocusing to conceptual building. *Computers in Human Behavior Reports*, *7*, 100223.
49. Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134.
50. Lin, P. Y., Chai, C. S., Jong, M. S. Y., Dai, Y., Guo, Y., & Qin, J. (2021). Modeling the structural relationship among primary students' motivation to learn artificial intelligence. *Computers and Education: Artificial Intelligence*, *2*, 100006.
51. Lin, T. B., Li, J. Y., Deng, F., & Lee, L. (2013). Understanding new media literacy: An explorative theoretical framework. *Journal of Educational Technology & Society*, *16*(4), 160–170.
52. List, A. (2019). Defining digital literacy development: An examination of pre-service teachers' beliefs. *Computers & Education*, *138*, 146–158.
53. Livingstone, S., & Helsper, E. (2010). Balancing opportunities and risks in teenagers' use of the internet: The role of online skills and internet self-efficacy. *New Media & Society*, *12*(2), 309–329.
54. Lo, L. S. (2023). The art and science of prompt engineering: A new literacy in the information age. *Internet Reference Services Quarterly*, *27*(4), 203–210.
55. Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, *74*(5), 570–581.
56. MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99.
57. Namkoong, J. E., & Henderson, M. D. (2016). Wanting a bird's eye to understand why: Motivated abstraction and causal uncertainty. *Journal of Experimental Social Psychology*, *64*, 57–71.
58. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041.
59. Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, *55*(3), 1082–1104.
60. Nimon, K., Zigarmi, D., Houson, D., Witt, D., & Diehl, J. (2011). The work cognition inventory: Initial evidence of construct validity. *Human Resource Development Quarterly*, *22*(1), 7–35.
61. Nunally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.
62. Orru, G., Piarulli, A., Conversano, C., & Gemignani, A. (2023). Human-like problem-solving abilities in large language models using ChatGPT. *Frontiers in Artificial Intelligence*, *6*, 1199350.
63. Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50.
64. Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780.
65. Perry, K. H. (2012). What is literacy?—A critical overview of sociocultural perspectives. *Journal of Language and Literacy Education*, *8*(1), 50–71.
66. Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98–125.
67. R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
68. Rajam, V., Reddy, A. B., & Banerjee, S. (2021). Explaining caste-based digital divide in India. *Telematics and*

- Informatics*, 65, 101719.
69. Revelle, W. (2018). Psych: Procedures for psychological, psychometric, and personality research. *R Package Version*, 1(10).
70. Ritzer, G., & Jurgenson, N. (2010). Production, consumption, prosumption: The nature of capitalism in the age of the digital 'prosumer'. *Journal of Consumer Culture*, 10(1), 13–36.
71. Robinson, S. C. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society*, 63, 101421.
72. Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
73. Saranto, K., & Hovenga, E. J. (2004). Information literacy—What it is about?: Literature review of the concept and the context. *International Journal of Medical Informatics*, 73(6), 503–513.
74. Simonsohn, U. (2009). Direct risk aversion: Evidence from risky prospects valued below their worst outcome. *Psychological Science*, 20(6), 686–692.
75. Southworth, J., Migliaccio, K., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, 4, 100127.
76. Staksrud, E., Olafsson, K., & Livingstone, S. (2013). Does the use of social networking sites increase children's risk of harm? *Computers in Human Behavior*, 29(1), 40–50.
77. Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25(2), 173–180.
78. Tan, A. S., Lee, C. J., & Chae, J. (2015). Exposure to health (mis) information: Lagged effects on young adults' health behaviors and potential pathways. *Journal of Communication*, 65(4), 674–698.
79. Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*, 1, 31–65.
80. van der Zeeuw, A., Van Deursen, A. J., & Jansen, G. (2019). Inequalities in the social use of the Internet of things: A capital and skills perspective. *New Media & Society*, 21(6), 1344–1361.
81. van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, 72, 577–588.
82. Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178.
83. Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21(1), 15.
84. Wang, B., Rau, P. L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 42(9), 1324–1337.
85. Wang, C. H., & Wu, C. L. (2022). Bridging the digital divide: The smart TV as a platform for digital literacy among the elderly. *Behaviour & Information Technology*, 41(12), 2546–2559.
86. Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2), 1–37.
87. Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management (JDM)*, 30(1), 61–79.
88. Wang, Y. Y., & Wang, Y. S. (2022). Development and validation of an artificial intelligence anxiety scale: An initial application in predicting motivated learning behavior. *Interactive Learning Environments*, 30(4), 619–634.
89. White, A., Breazeale, M., & Collier, J. E. (2012). The effects of perceived fairness on customer responses to retailer SST push policies. *Journal of Retailing*, 88(2), 250–261.

90. Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 1–27.

[URL]

91. 방송통신위원회 (2024). **2023년 지능정보사회 이용자 패널조사**.
<https://www.kcc.go.kr/user.do;jsessionid=IAQBoWhhXJkTOg9Zt08sadcAu2FS8kBEgfvfp7Zc.servlet-aihgcldhome10?mode=view&page=A02060400&dc=K02060400&boardId=1030&cp=1&boardSeq=61938>
92. OpenAI (2024). ChatGPT: An AI language model [Language model]. *OpenAI*. <https://www.openai.com/research>

● 저 자 소 개 ●



노 환 호 (Hwan-Ho Noh)

현재 연세대학교 바른ICT연구소 연구교수로 재직 중이다. 광운대학교에서 심리학 박사 학위를 취득하였다. 주요 연구 분야는 사회적 요인에 따른 의사결정 연구, 소셜미디어 광고 효과, 인공지능의 사회적 역할과 신뢰 등이다. 지금까지 Journal of Consumer Behaviour, Behaviour & Information Technology, 소비자학연구 등에 논문을 발표했다.



김 현 정 (Hyeonjeong Kim)

현재 연세대학교 바른ICT연구소 연구교수로 재직 중이다. 연세대학교 정보대학원에서 정보시스템박사 학위를 취득하였다. 주요 관심분야는 메타버스, 머신러닝 및 딥러닝, 이미지 객체 탐지, 악성댓글 탐지 및 디지털 리터러시 등이다. 지금까지 Technology Forecasting and Social Change, 지식경영연구, 디지털콘텐츠학회 논문지 등 학술지에 논문을 발표하였다.



김 민 진 (Minjin Kim)

현재 연세대학교 바른ICT연구소 연구교수로 재직 중이다. University of East Anglia에서 국제개발학 박사 학위를 취득하였다. 주요 연구 분야는 정보격차해소, ICT for development, FinTech, Financial development 등이다. 지금까지 Information Technology for Development, 소비자학연구, 정치정보연구 등에 논문을 발표했다.

〈 Abstract 〉

Development and Validation of a Korean Generative AI Literacy Scale

Hwan-Ho Noh^{*}, Hyeonjeong Kim^{**}, Minjin Kim^{***}

Literacy initially referred to the ability to read and understand written documents and processed information. With the advancement of digital technology, the scope of literacy expanded to include the access and use of digital information, evolving into the concept of digital literacy. The application and purpose of digital literacy vary across different fields, leading to the use of various terminologies. This study focuses on generative artificial intelligence (AI), which is gaining increasing importance in the AI era, to assess users' literacy levels. The research aimed to extend the concept of literacy proposed in previous studies and develop a tool suitable for Korean users. Through exploratory factor analysis, we identified that generative AI literacy consists of four factors: AI utilization ability, critical evaluation, ethical use, and creative application. Subsequently, confirmatory factor analysis validated the statistical appropriateness of the model structure composed of these four factors. Additionally, correlation analyses between the newly developed literacy tool and existing AI literacy scales and AI service evaluation tools revealed significant relationships, confirming the validity of the tool. Finally, the implications, limitations, and directions for future research are discussed.

Key words: Generative AI literacy, Measurement tool development, Validity assessment, Ethics, Use of information technology

* Barun ICT Research Center, Yonsei University

** Barun ICT Research Center, Yonsei University

*** Barun ICT Research Center, Yonsei University