한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2025. Vol. 38, No. 1, 1-26
https://doi.org/10.24230/kjiop.v38i1.1-26

인공지능 기반의 부정적 피드백에 대한 수용도: 위협과 신뢰성의 매개 효과를 중심으로

전 세 희　인 성 빈　강 지 원　박 지 영†

덕성여자대학교 심리학과

인공지능 기술 개발이 가속화됨에 따라 일터에서 인공지능 기술 기반의 피드백을 받는 기회가 많아지고 있다. 본 연구에서는 창의적 수행 과제를 수행한 후 인공지능 피드백에 대한 사람들의 반응과 태도를 살펴보았다. 피드백 수용 이론을 기반으로 인공지능 기술 기반의 부정적 피드백을 받은 경우 사람이 제공하는 부정적 피드백을 받을 때보다 수용도가 낮을 것이라고 예상하였으며, 이는 정체성 위협과 현실적 위협 수준이 높아지고, 피드백 원천에 대한 신뢰성이 낮아지기 때문일 것이라고 예상하였다. 또 두 위협의 종류가 피드백 수용도에 미치는 영향을 내적 통제소재가 조절할 것이라고 예상하였다. 실험 연구 결과, 인공지능 기반의 부정적 피드백을 받은 경우 사람이 제공하는 부정적 피드백을 받을 때보다 정체성 위협과 현실적 위협을 많이 느꼈다. 예상과 달리, 두 조건 간 피드백 수용도 및 신뢰도의 차이는 유의하지 않았다. 매개 효과 검증 결과, 정체성 위협을 많이 느낄수록 인공지능 피드백에 대한 수용도는 낮아졌으나, 가설과 달리 현실적 위협 수준이 높아질수록 인공지능 피드백에 대한 수용도는 높아졌다. 조절 효과 검증 결과, 정체성 위협과 인공지능 피드백 수용도의 관계를 내적 통제소재가 조절하는 것을 확인하였다. 본 연구 결과를 바탕으로 연구의 의의, 한계점, 후속 연구 방향을 논의하였다.

주요어 : 인공지능, 피드백 수용도, 신뢰성, 정체성 위협, 현실적 위협, 내적 통제소재

† 교신저자 : 박지영, 덕성여자대학교 심리학과, 서울특별시 삼양로 144길 33
E-mail: jiyoungpark@duksung.ac.kr

Copyright ⓒ 2025, The Korean Society for Industrial and Organizational Psychology. This is an open-access article distributed under the terms of the Creative Commons Attribution –NonCommercial Licenses (https://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

최근 인공지능 기술의 발전 및 확산으로 조직 및 개인의 삶에 인공지능 기술이 다양한 형태로 적용되고 있다(Kaplan & Haenlein, 2019; Tong et al., 2021). 인공지능을 '새로운 전기(new electricity)'라고 부르기도 하는데(Ng, 2018), 이는 전기가 인류에게 미친 영향처럼 인공지능 기술의 영향력이 넓다는 것을 비유하는 것이다. 인공지능 기술의 영향력이 지배적인 이유는 다른 기술과 달리, 인공지능은 자기 학습이 가능하고 목적에 맞게 결과물을 산출할 수 있기 때문이다(Kaplan & Haenlein, 2019). 특히, 인공지능 기술은 개별화된 상호작용을 기반으로 즉각적이고 다양한 피드백을 제공할 수 있으며, 이러한 특징으로 인해 일터와 일상에서 인공지능 피드백의 활용이 증가하고 있다(문혜진, 남상희, 2022; Jia et al., 2024; Tong et al., 2021).

피드백 연구에 따르면 동일한 피드백을 받더라도 피드백 수용도는 사람마다 다르며, 피드백 수용도는 피드백을 실질적으로 사용하고 수행을 향상시키는 데 영향을 미친다(Anseel & Lievens, 2009). 피드백 수용도에 영향을 미치는 주요한 요인 중 하나는 피드백의 원천(feedback source)이다. 즉, 누가 피드백을 제공했느냐에 따라 동일한 내용의 피드백이라고 할지라도 수용자는 정서적, 인지적으로 다르게 반응하고 피드백 수용도 또한 달라진다(Brett et al., 2001; Kluger & DeNisi, 1996). 인공지능 기술이 피드백의 원천이 됨에 따라, 인공지능 기술이 제공하는 피드백을 사람이 제공하는 피드백과 다르게 평가하는지, 다르다면 차이의 이유를 파악하는 것이 필요하다.

사람들은 인공지능 기술 기반의 피드백을 사람의 피드백과 다르게 평가한다. 서구권에서 진행된 연구에서는 인공지능의 의사 결정 및 결과물을 사람의 것보다 부정적으로 수용하는 경향이 있다(Magni et al., 2024; Nargesi, 2021). 하지만, 인공지능 기술의 피드백에 대한 수용도 및 태도는 인공지능 기술이 적용된 영역과 문화에 따라 다르기도 하다(Folk et al., 2023; Mantello et al., 2023). 예를 들어, 한국을 비롯한 동아시아권 사람들은 미국 및 서구권보다 인공지능에 대해서 더 긍정적으로 평가하는 경향이 있다(Folk et al., 2023; Kim & Kim, 2022). 또, 영역에 따라서도 다른데 국내 성인을 대상으로 한 연구에 따르면 군대와 같이 자신의 삶에 큰 영향을 미치고 자신과 깊게 관여되어 있는 상황에서는 인공지능 기술이 의사결정을 내리는 것을 꺼리는 것으로 나타났다(도은영 외, 2021). 이런 연구 결과들로 인해, 인공지능에 대한 태도를 연구할 때 인공지능 기술이라는 큰 덩어리가 아니라 구체적인 인공지능 기술의 특징 혹은 적용되는 과업에 맞춰 인공지능 기술에 대한 사람들의 반응 및 태도를 살펴볼 필요가 있다고 학자들은 주장한다(Landers & Marin, 2021).

이와 같은 주장에 근간하여 본 연구에서는 창의적 수행 과제에 대한 인공지능 기술의 피드백에 초점을 두었다. 기존 연구에 따르면 감정이나 창의성과 관련된 분야에서 인공지능 기술에 대한 부정적인 반응이 더욱 두드러진다(Magni et al., 2024). 사람들은 인간의 고유한 영역으로 간주되는 창의성, 추상적 생각, 정서와 관련된 과업에 인공지능이 적용될 경우 부정적인 반응을 보인다(Kim & Duhachek, 2020; Magni et al., 2024). 한국인 역시, 인간의 고유성과 관련된 인공지능 기술에 대해서는 심리적 저항감이 높은 것으로 나타났다(도은영 외, 2021). 부정적인 반응에도 불구하고 인공지능 기술이 시, 음악 작곡과 같은 창의적 영역에

적용하는 사례는 증가하고 있다(du Sautoy, 2019; Gervás, 2019). 또한 인공지능 기술을 이용한 창의적 작품은 상업적으로도 성과를 보이고 있다. 2018년 크리스티(Christie's) 경매에서 인공지능이 제작한 작품이 처음으로 경매에 나와 높은 가격에 팔렸으며(Cohn, 2018), 밀라노 디자인 위크에서 인공지능 기술로 디자인하고 3D로 프린팅된 의자가 처음으로 전시되기도 하여(노동균, 2019), 인공지능 기술이 창의적 결과물을 내는데 유용하게 사용될 수 있음을 알 수 있다. 일상에서도 생성형 인공지능 프로그램을 사용하여 창의적 작품 및 창의적 과업을 수행할 수 있게 됨에 따라 창의적 수행에 대한 인공지능 기술의 피드백 수용에 대한 연구가 필요한 시점이다. 하지만 관련 연구는 부족하고, 특히 국내 성인을 대상으로 한 연구는 찾아보기 힘들다.

본 연구에서는 국내 성인을 대상으로 창의적 과제 수행 시, 인공지능과 사람의 피드백에 대한 피드백의 수용도의 차이와 그 이유에 대해서 살펴보고자 한다. 피드백 수용 이론을 근간으로(Ilgen et al., 1979) 피드백 원천에 대한 신뢰성과 정서적 반응이 피드백 수용도에 영향을 줄 것이라고 가정하였고, 정서적 반응 중 위협 수준의 차이를 확인하고자 한다. 인공지능에 대한 위협이 현실적 위협과 정체성 위협으로 나뉜다는 점을 고려하여(Jackson et al., 2020) 본 연구에서도 두 종류의 위협에 초점을 두었다. 두 종류의 위협과 신뢰성의 다중의 메커니즘을 확인하여 인공지능 기술에 대한 피드백 수용도에 미치는 인지적, 정서적 메커니즘을 파악하고자 한다. 추가로, 위협이 피드백 수용도에 미치는 영향은 개인의 성격에 따라 달라질 수 있기 때문에(Ilgen et al., 1979) 내적 통제소재(internal locus of control)의 조절 효과를 확인하고자 한다.

창의적 과제에 대한 인공지능의 부정적 피드백에 대한 수용도

Ilgen 등(1979)은 피드백 수용도에 영향을 끼치는 주요 요인을 피드백의 내용(message), 피드백 원천(source), 수용자(recipient)로 나누어 설명하였다. 피드백의 내용은 유인가(valence)에 따른 분류로서 긍정적 피드백이 부정적 피드백보다 수용도가 높다는 것은 많은 연구에서 확인된 바 있다(Ilgen, 1971; Kluger & DeNisi, 1996). 부정적 피드백에 대해 수용도가 낮은 이유는 자기(self) 개념과 관련지어 설명할 수 있다. 긍정적 피드백은 자신의 긍정적 자기를 강화하기 때문에 수용자가 해당 피드백이 정확하다고 인식한다. 반면 부정적 피드백의 경우, 자기 개념에 부정적인 영향을 줄 수 있으므로 자기에 대한 부정적인 정보와 지식을 받아들이기 원하지 않으며 해당 지식과 정보가 부정확하다고 판단하여 거부한다(Crocker & Wolfe, 2001; Sherman & Cohen, 2002). 이런 연구 결과를 기반으로 본 연구에서는 수용도가 낮은 부정적 피드백에 초점을 두고 인공지능 기술 기반의 부정적 피드백의 수용도에 영향을 주는 요인을 살펴보고자 하였다.

창의적 과제를 수행한 후 동일한 부정적 피드백이라고 해도 인공지능 기술이 제공할 경우 사람이 제공하는 경우보다 피드백 수용도가 낮을 것이라고 예상하였다. 인공지능 및 사람에 대해 가지고 있는 기존의 지식, 편향, 믿음이 해당 피드백을 해석하는데 영향을 줄 수 있다(Magni et al., 2024; Kim & Duhachek, 2020). 사람들은 인공지능 기술이 인간의 영역이라고 믿어 온 창의적인 분야에서 피드백을

제공할 경우 인공지능 기술이 이러한 추상적인 생각을 하기 힘들다고 인지하며, 창의적인 결과물을 산출하는데 노력을 덜 기울인다고 믿는다(Magni et al., 2024; Kim & Duhachek, 2020). 인공지능 기술에 대해 지니고 있는 이러한 믿음, 편향이 피드백 수용도에 영향을 줄 수 있다. 실제로, 인공지능 기술에 대한 사람들의 오해를 선제적으로 설명하여 인공지능도 정서적인 영역에서 전문성을 가지고 있음을 설명하거나(Longoni & Cian, 2022) 창의적인 작품을 만들기 위해 인공지능 기술도 실제로 노력한다는 것을 보여줄 경우(Chamberlain et al., 2018), 인공지능에 대한 편향이나 부정적인 태도가 줄어들기도 한다.

창의적 과제의 경우 인공지능 기술에 대한 편향을 지니고 있고, 사람의 영역이라는 믿음이 있기 때문에 이러한 믿음이 정서적, 인지적인 반응에 영향을 주고 이것이 피드백을 덜 수용하게 만드는 요인이 될 것이다. 특히, 인공지능 기술에 대해 지니고 있는 신뢰 수준이 영향을 줄 수 있다. 동일한 결과물이라고 해도 사람이 아닌 인공지능 기술에 의해 도출된 결과물이라고 설명할 경우, 결과물 및 절차에 대한 신뢰 수준(Nargesi, 2021), 공정성 인식(Bedemariam & Wessel, 2023)이 낮고 이는 수용도를 낮추는 요인이 될 것이다. 예를 들어, Nargesi(2021)의 연구에서 참가자들에게 동일한 피드백을 제공하면서 한 그룹에서는 평가자에 대한 정보를 사람이라 하였고 다른 그룹에게는 인공지능의 피드백이라고 설명하였다. 사람 평가자의 피드백을 받은 사람들은 인공지능의 피드백이라고 믿은 사람보다 피드백을 더 신뢰하고 더 수용하였다. 채용 과정에서도 부정적인 평가를 받았을 때, 인공지능의 평가를 사람의 평가보다 덜 공정하다고 인식하였다(Bedemariam & Wessel, 2023). 이와 같이 인공지능 기술에 대해서 사람들이 지각하는 낮은 공정성과 신뢰 수준이 피드백 수용도를 낮추는 요인이 될 것이다.

가설 1. 인공지능의 부정적 피드백 조건에서 사람의 부정적 피드백 조건보다 더 낮은 피드백 수용도를 보일 것이다.

현실적 위협과 정체성 위협의 매개 효과

인공지능 기술 기반의 부정적 피드백은 사람의 부정적 피드백보다 위협을 높이는 요인이 될 것이다. 개인은 여러 사회적 집단에 속하게 되며 자신이 소속된 집단을 내집단, 그렇지 않은 집단을 외집단으로 구별 지어 생각한다(Hewstone, et al., 2002; Riek et al., 2006). 일상 생활에서는 인종, 나이, 성별과 같은 특성이 내집단과 외집단을 구별짓는 주요한 요소이지만, 이러한 요인을 넘어서는 현저한 표식(salient markers)이 등장하면 다른 인종, 다른 성별이라고 해도 하나의 내집단으로 인식하며 새로운 범주화가 일어나기도 한다(Jackson et al., 2020). 인공지능 기술이라는 현저한 표식이 등장할 경우 사람들은 인간을 닮은 기계라는 특성에 주목하고 인간과 인간이 아닌 집단으로 범주화하여 인간을 내집단, 인간이 아닌 인공지능 기술을 외집단으로 분류한다. 한 연구에 따르면 인공지능 기술 혹은 자동화 로봇으로 변화하는 사회에 대한 기사를 읽은 사람들은 인간과 인간이 아닌 특성에 집중하여 일반적으로 외집단으로 분류될 수 있는 이민자들을 '인간'이라는 범주 안에 속하는 내집단으로 간주한다(Jackson et al., 2020). 즉 인공지능 기술은 현저한 표식으로 작동하여 인간을 내

집단으로 인공지능 기술을 외집단으로 범주화 할 수 있다(Jackson et al., 2020). 이처럼 외집단으로 분류된 집단에 대해서 사람들은 현실적 위협(realistic threat)과 정체성 위협(identity threat)을 느낀다(Jackson et al., 2020; Riek et al., 2006). 인공지능이 야기하는 현실적 위협은 사람의 자원, 안전 및 육체적 안녕을 위협을 의미하며(Riek et al., 2006), 정체성 위협은 사람의 정체성 또는 독창성에 대한 위협을 의미한다(Hewstone, et al., 2002; Riek et al., 2006).

창의적 수행 과제에 대해서 인공지능 기술 기반의 부정적 피드백을 받을 경우 현실적 위협이 높아질 것이라고 예상하였다. 사람들은 자신에 대해 긍정적인 자기 개념을 가지고 있으며 이를 유지하고자 한다(Heine & Lehman, 1997). 사람들은 자신의 수행이 실패하거나 수행에 대해 부정적 피드백을 받으면 자기가 위협받았다고 인식한다(Crocker & Wolfe, 2001). 특히, 성과에 대한 부정적인 평가는 특정 과제 수행이라는 목표를 달성하지 못했다는 인식을 주며 목표 달성의 실패는 부정적인 감정을 유발하는 주요 요인이 된다(Kluger & DeNisi, 1996). 자신이 수행한 창의적 과제에 대해서 인공지능 기반의 부정적인 피드백을 받을 경우, 인공지능의 능력에 대해서 자각하게 되고 이는 자신의 직업을 대체할 수 있는 존재로 인식하여 현실적 위협이 높아질 것이다. 인공지능 기술이 자신의 일을 대체한다는 직접적인 정보가 주어지지 않아도 인공지능 기술의 유능함이나 자율적인 특성을 지각하면 이는 현실적 위협을 높일 수 있다. 예를 들어, 인공지능 기술의 글쓰기 능력이 우수하다는 기사를 읽거나(박지영, 정예슬, 2020), 자동화 로봇이 청소하거나 노인을 돌보는 영상을 보는 상황에서도 사람들은 현실적 위협을 느꼈다(Złotowski et al., 2017). 창의적 수행 과제에 대한 피드백은 창의성에 대한 이해 및 의사결정 능력과 같은 자율적인 능력이 요구되기 때문에 인공지능 기술 기반의 부정적 피드백을 받을 경우 인공지능 기술의 특성에 대해 주목하게 되고 이는 현실적 위협을 높이는 요인이 될 것이다.

인공지능이 창의적 과제에 대해서 부정적 피드백을 제공할 경우, 정체성 위협 수준도 증가할 것이다. 사람들은 자신이 속한 내집단을 특별하게 생각하고 외집단과는 다르게 보려는 강한 욕구를 가지고 있으며, 정체성 위협은 이러한 집단의 독특성을 유지하려는 욕구와 관련되어 있다(Hewstone et al., 2002). 내집단의 독특성에 대한 느낌이 사라지면 외집단과 구별되지 않는다고 느끼며 자신의 정체성에 대한 위협을 느끼게 된다(Hewstone et al., 2002; Jackson et al., 2020). 실제로 인공지능이 자율성과 같은 인간의 속성을 지니고, 인간과 유사하다고 인식될 경우, 기계와 사람 사이의 경계를 흐릴 수 있다고 생각하여 정체성 위협을 느꼈다(Złotowski et al., 2017). 특히, 사람들은 창의성을 인간의 고유한 능력으로 인식하기도 하며, 인공지능이 상업적 영역보다 추상적 그림과 같은 창의적인 영역에서 적용될 때 더 부정적인 태도를 보였다(Magni et al., 2024). 이러한 연구 결과를 바탕으로 인공지능이 창의적 수행 과제에 대해서 부정적인 피드백을 제공할 경우, 사람이 제공할 때보다 정체성 위협을 더 많이 느낄 것이라고 예상하였다.

가설 2. 인공지능의 부정적 피드백 조건에서 사람의 부정적 피드백 조건보다 현실적 위협(가설 2a)과 정체성 위협(가설 2b)을 더 많이 느낄 것이다.

이렇게 증가한 현실적 위협과 정체성 위협은 인공지능의 피드백에 대한 피드백 수용도를 낮추는 요인이 될 것이다. 자신이 가진 능력이나 자원에 대해서 부정적인 피드백을 받을 때 사람들은 자기를 보호하고 가치를 강화하고자 하며, 이는 다양한 인지적 반응과 행동적 반응을 불러일으킨다(Heine & Lehman, 1997). 자기를 보호하기 위해 사람들이 취하는 대처 방식 중 하나는 위협을 무시하고 피해버리거나 위협을 가하는 정보를 수용하지 않는 방어적 편향(defensive biases)을 갖는 것이다(Sherman & Cohen, 2002). 인공지능으로부터 부정적 피드백을 받았을 때 위협을 느끼고 방어적 편향을 통해 해당 피드백은 이상적인 자기와 맞지 않다고 판단하여 수용도가 낮아질 수 있다. 또, 위협은 회피 경향성을 불러일으키며, 새로운 것을 수용하지 않는 태도를 촉발시킨다(Lerner & Keltner, 2001). 실제로, 현실적 위협 및 정체성 위협은 인공지능 기반의 제안을 수용하지 않거나, 인공지능에 대한 전반적인 부정적인 태도를 야기하는 요인이 되었다(Jackson et al., 2020; Złotowski et al., 2017). 이러한 연구를 바탕으로 현실적 위협과 정체성 위협이 피드백 수용도를 낮추는 요인이 될 것이라고 예상하였다.

가설 3. 인공지능 기술의 부정적 피드백이 피드백 수용도에 미치는 영향은 현실적 위협(가설 3a) 및 정체성 위협(가설 3b)에 의해 매개될 것이다.

피드백 원천에 대한 신뢰성의 매개 효과

위협과 함께, 피드백 제공자가 인공지능 기술인지 사람인지에 따라 피드백 원천에 대한 신뢰성(feedback source credibility)이 차이가 날 것이라고 예상하였다. 피드백 원천의 신뢰성이란 피드백의 원천이 평가하고자 하는 행동이나 과제에 대해 정확하게 판단할 수 있는 전문성을 지녔다고 믿는 정도 및 피드백 원천 자체를 믿을 수 있는 존재라고 인식하는 정도를 뜻한다(Ilgen et al., 1979). 피드백 수신자는 피드백을 수용할 때 메시지와 함께 피드백의 원천을 파악하며, 피드백 원천에 대한 신뢰성은 피드백 수용도를 향상시킨다(Ilgen et al., 1979; Reinke, 2003). 지각된 원천의 신뢰성이 높다는 것은 피드백 수신자가 피드백의 제공자를 믿고 제공자가 해당 분야에 충분한 전문성을 가지고 있다는 인식한다는 것을 뜻한다(Ilgen et al., 1979).

창의성은 인간의 고유한 영역이라고 믿기 때문에 인공지능 기술보다 사람의 특성을 이해하는 사람의 피드백을 더 신뢰할 것이다. 부정적 피드백은 자기 개념에 부정적인 영향을 주는 것으로 수용하기 어렵고 이에 피드백 원천의 중요성이 더욱 부각된다(Bannister, 1986). 신뢰는 위험과 불확실한 상황에서 기대한 것을 수행하는 데 필요한 속성을 가지고 있다는 믿음을 뜻하며, 인공지능 기술은 사람이 통제할 수 없는 부분이 있기 때문에 신뢰의 중요성이 더욱 부각된다(Glikson & Woolley, 2020). 인공지능 기술의 자율적인 특성으로 인공지능 기술의 산출물에는 설명이 불가능한 측면이 존재한다(Kaplan & Haenlein, 2019). 인공지능 기술 기반의 부정적 피드백을 받을 경우, 왜 이러한 산출물이 도출되었는지에 대한 의구심을 갖고 자신의 통제감을 지키고자 사람보다 인공지능 기술의 피드백을 더 신뢰하지 않는 결과로 이어질 것이다. 실제로, Nargesi(2021)의 연구에서 사람과 인공지능의

피드백을 비교했을 때 사람의 피드백을 더 신뢰하며, 객관적인 피드백보다 주관적인 피드백일 때 사람과 인공지능의 피드백 원천의 신뢰성의 차이가 더 크다는 것을 확인한 바 있다. 이러한 연구 결과를 기반으로, 동일한 부정적인 피드백을 제공하더라도 인공지능 기반의 피드백에 대해서 사람보다 더 낮은 신뢰성을 보일 것이라고 예측하였다.

가설 4. 인공지능의 부정적 피드백 조건에서 사람의 부정적 피드백 조건보다 더 낮은 신뢰성을 지각할 것이다.

피드백 원천에 대한 신뢰성 수준은 피드백 수용도에 영향을 미칠 것이다. 피드백 수용 이론에 따르면 피드백 원천에 대한 신뢰가 높을수록 피드백을 더 수용한다(Albright & Levy, 1995; Bannister, 1986; Ilgen et al., 1979). 사람들은 일반적으로 자신과 가깝고 직접적인 교류를 한 사람들이나 내집단이라고 생각하는 평가자의 피드백을 더 신뢰하고 수용한다(Brett & Atwater, 2001). 사람들은 인공지능을 외집단이라고 판단하기 때문에(Jackson et al., 2020) 동일한 피드백이라고 해도 내집단이라고 느끼는 사람 평가자가 제공하는 피드백을 더 신뢰하고 이러한 신뢰성의 차이는 피드백 수용도의 차이로 이어질 것이다. 한 연구에서 미국 성인들은 인공지능의 피드백을 사람의 피드백보다 덜 수용했는데 이는 피드백 원천에 대한 신뢰성이 낮기 때문인 것으로 나타났다(Nargesi, 2021). 이러한 연구 결과를 기반으로, 원천에 대한 낮은 신뢰성이 인공지능 기반의 피드백 수용도를 낮추는 요인이 될 것이라고 예상하였다.

가설 5. 인공지능 기술의 부정적 피드백이 피드백 수용도에 미치는 영향은 지각된 원천의 신뢰성에 의해 매개될 것이다.

내적 통제소재의 조절 효과

인공지능이 제공하는 부정적 피드백은 현실적 위협과 정체성 위협을 불러일으키고 이는 피드백 수용도를 낮추지만 개인의 성격 특성에 따라 위협이 피드백 수용도에 미치는 영향이 달라질 수 있다. 통제소재(locus of control)는 개인의 느낌, 성취나 실패 또는 행동 결과를 설명하는 인과적 기제 중 하나로, 자신의 행동이나 감정을 지배하는 원인을 자신의 내부 또는 외부에 두는지에 따라 내적 통제소재와 외적 통제소재로 나뉜다(Rotter, 1966). 자신의 행동과 이에 따르는 결과를 스스로 통제할 수 있다고 생각할수록 내적 통제소재를 지니며, 자신의 행동과 행동의 결과가 자신과는 무관한 외적인 요인에 의해 지배된다고 생각하는 경우 외적 통제소재를 지닌다고 설명한다(Rotter, 1966).

내적 통제소재가 높은 사람은 인공지능 기반의 부정적 피드백에 대해서 위협을 느끼더라도 피드백을 통해 배울 점이 있다고 생각하며 건설적으로 반응할 것이다. 또한 학습 및 성과 향상에 대한 동기를 지니고 있으며, 문제 해결을 성공적으로 할 수 있다고 믿는다(Rotter, 1966). 따라서 부정적 피드백으로 인해 위협을 느낄 때에도 적극적으로 수용하고자 할 것이다. 인공지능 기술 피드백에 대해 내적 통제소재의 역할을 직접적으로 살펴본 연구는 없지만, 새로운 기술을 적극적으로 수용하거나 과업에 대해 높은 목표를 지니고 학습하는 경향이 있는 사람들은 인공지능 기술에

대해 부정적인 정서를 느끼더라도 인공지능의 기능적 측면에 대해서 긍정적으로 평가했다(Park & Woo, 2024). 또, 내적 통제소재가 높은 사람들은 부정적인 결과에 대해서도 자신의 노력이나 학습을 통해 개선할 수 있다고 인식하며 도전적인 상황에서 문제 중심적인 대처를 하는 것으로 나타났다(Ng et al., 2006). 이러한 연구 결과를 기반으로 내적 통제소재는 인공지능이 제공한 부정적 피드백에 대해 느끼는 정체성, 현실적 위협과 피드백 수용도의 부적 관계를 완화하는 역할을 할 것으로 예측하였다.

가설 6. 내적 통제소재는 인공지능이 제공하는 부정적 피드백에 대한 현실적 위협(가설 6a)과 정체성 위협(가설 6b)이 피드백 수용도에 미치는 부적 영향력을 약화시킬 것이다.

연구대상 및 절차

가설을 검증하기 위해 동일한 참가자를 대상으로 두 차례에 걸쳐 설문을 진행하였다. 첫 번째 설문에서는 두 종류의 창의적 수행 과제를 진행한 후, 본인이 지각한 수행 수준이 피드백 수용도에 영향을 준다는 연구 결과를 기반으로(Vance & Colella, 1990) 수행한 과제에 대해서 자신들이 생각하는 수행 수준을 평가하도록 하였다. 이후 내적 통제소재, 인구통계학적 변인에 대해 응답하도록 하였다. 두 번째 설문은 첫 번째 설문을 완료한 사람들을 대상으로 24시간 후 발송되었으며 두 번째 설문이 발송되고 48시간 이내에 응답한 답변만 유효한 것으로 간주하였다. 두 번째 설문에서는 인공지능 피드백 조건과 사람 피드백 조건으로 나뉘어 두 조건에 참가자를 무선 할당하였다. 각 조건 별로 인공지능 혹은 사람이 피드백을 준다고 명시하고, 과제에 대한 부정적 피드백을 제공하였다. 참가자들은 피드백을 읽은 후, 피드백을 받고 난 후 느끼는 위협 수준, 피드백 제공자에 대한 신뢰성, 피드백 수용도 수준을 응답하였다.

본 연구에 앞서 해당 조작이 효과적인지 확인하기 위해 예비 연구를 진행하였다. 예비 연구는 편의 표집 방식을 통해 참가자들을 수집하였다. 총 32명의 참가자를 수집하였으며, 예비 연구에 참여한 사람들은 대부분 20대였으며(93.75%) 성별은 여성이 59.37%, 남성이 40.62%로 나타났다. 두 번째 설문까지 기간 내에 모두 완료한 사람은 32명 중 22명(68.5%)이었다. 인공지능 조건과 사람 조건에서 각각 11명이 설문을 완료하였다. 조작 점검을 확인한 결과, 인공지능 조건에서는 전체 11명 모두가 인공지능이 피드백 제공자라고 응답하였고, 사람 조건에서는 11명 중 8명(72.73%)이 사람이 피드백 제공자라고 인식하였다. 참가자들이 첫 번째 설문에서 두 가지 창의적 과제를 성실히 응답하였고 두 번째 설문을 통해 조작은 효과적이라고 판단하여 본 연구를 진행하기로 결정하였다.

본 연구는 국내 온라인 설문 업체를 통해 참가자를 수집하였다. 연구에 참가하기 전 연구 목적과 절차를 안내하였다. 만 20세 이상에서 40세 미만의 한국인 300명을 수집하고자 하였다. 두 개의 설문 간 시간 차이로 인한 결측 자료와 불성실 응답자를 고려하여 약 30%를 더 수집하여 첫 번째 설문에서는 총 398명의 참여자의 자료를 수집하였다. 첫 번째 설문에 응답한 사람들을 대상으로 두 번째 설문을 24시간 이후 발송하였고, 48시간 이내 응답한 사람의 답변을 포함하였다. 두 번째

설문에 응답한 398명 중 319명이 응답하였고 80.15%의 응답률을 보였다.

두 번째 설문에서는 주의 확인 문항(“본 질문에 ‘매우 그렇다’를 응답해 주시길 바랍니다”) 두 개를 포함하였고, 두 문항을 모두 잘못 응답한 21명을 제외하였다. 실험이 효과적으로 진행되었는지 확인하기 위해 “귀하의 창의적 과제를 평가한 주체는 누구입니까?”와 같은 조작 확인 문항을 사용하였고 인공지능 혹은 사람 전문가 둘 중 한 개의 항목에 응답하도록 하였다. 조작 확인 문항에서 총 68명의 참가자들이 조건과는 다른 응답을 제공하였다. 52명의 참가자들이 사람 피드백 조건이었음에도 ‘인공지능’이 피드백을 제공했다고 응답하였으며, 인공지능 조건에서도 ‘사람’이 피드백을 제공했다고 응답하기도(n = 16명) 했다. 이는 인공지능 기술이 다양한 과업 및 환경에서 사용됨에 따라 참가자들이 조건에 관계없이 자신들의 생각이나 판단을 기반으로 응답한 것으로 유추되며, 사람이 응답할 경우에도 인공지능 기술의 도움을 받거나 인공지능을 사용했을 것이라고 판단한 것으로 추측한다. 본 연구 뿐 아니라 예비 연구에도 이러한 경향성이 나타났으며, 한국과 유럽 국가의 성인을 대상으로 한 다른 인공지능 연구에서도 이러한 현상이 나타난 바 있다(예: 김시내, 박지영, 2024; Magni et al., 2024). 조작 확인 문항에 잘못 응답한 참가자들에게는 실험 조작이 효과적으로 나타나지 않았다고 판단하여, 각 조건 별로 조작 질문에 바르게 응답한 참가자들의 자료만 포함하였다. 최종적으로 230명의 참가자의 자료가 본 연구의 분석에 포함되었다. 참가자들의 평균 나이는 29.73세(SD = 5.30)였고, 성별 분포는 여성 51.73%, 남성 48.26%이었으며, 참가자들의 교육 수준은 대학교 졸업 66.09%, 대학교 재학 17.83%, 고등학교 졸업 8.70%, 석사 이상 졸업자가 7.39%인 것으로 나타났다.

측정도구

창의적 수행 과제

창의적 수행 과제는 대체 용도 과제(alternative use task)와 창의적 상상력 과제(creative imagery task)와 같이 두 개의 과제를 사용하였다. 대체 용도 과제는 Guilford(1959)가 개발한 것으로 한 가지의 물건을 보고 본래의 쓰임이 아닌 다른 용도를 나열하는 것이다. 본 연구에서는 종이 클립을 사진으로 제시하고 기존 용도 외의 많은 용도를 나열하도록 하였다. 온라인으로 과제 수행 시, 이 분이 지나야 다음 화면으로 넘어갈 수 있도록 설정하였다. 종이 클립의 다른 용도에 대해서 참가자들은 ‘자물쇠 따기’, ‘여러 개의 종이 클립을 이어 줄자처럼 사용하기’와 같은 용도를 응답하였다.

두 번째 과제는 창의적 상상력 과제로 Chiu (2012)에서 사용한 과제를 사용하였다. 그림 1에 제시된 바와 같이 15개의 그림을 참가자들에게 보여주고, 세 개를 선택하여 가상의 물건을 발명하는 과제이다. 가구, 운송 기구, 장난감 등 여덟 개의 범주 중 한 개의 범주를 선택하여 해당 범주에 속하는 가상의 물건을 발명하도록 지시하는데, 본 연구에서는 장난감 범주 안에서 가상의 물건을 발명하도록 하였다. 참가자들은 세 개의 물건을 선택하고, 삼 분간 하나의 장난감을 발명하도록 하였다. 참가자들에게 선택한 세 개의 물건의 번호를 작성하고, 발명한 장난감의 이름과 장난감의 사용 방법을 작성하도록 설명하였다. 참가자



그림 1. 창의적 상상력 과제에서 사용한 그림

들의 예시 답변으로는 원뿔(4), 원기둥(5), 와이어(6)를 이용하여 '크리스마스 트리'라는 이름의 장난감을 만들었고, 장난감의 사용 방법을 '와이어를 당기면 노래가 나오는 크리스마스 트리'라고 설명한 답변이 있었다. 또 다른 참가자는 원기둥(5), 와이어(6), 튜브(7)를 이용하여 '롤링썬더 드래곤'이라는 이름의 장난감을 만들었고, 사용 방법을 '이 장난감은 둥근 원통 모양의 늘어났다 줄어들었다 하는 애완동물용 장난감이다(...) 애완동물이 이 통로를 지나갈 수 있으며 늘었다 줄었다 하기 때문에 길이를 늘릴 수도 있고 작게 만들어 다양한 장난감과 연계할 수 있다.' 등과 설명하였다.

부정적 피드백

부정적 피드백은 Fodor과 Carver(2000)의 연구에서 사용된 창의적 수행 과제에 대한 부정적 피드백의 형식을 참고하여 참가자가 평균 이하의 점수를 얻었고, 참가자가 과제 수행에 어려움을 겪었을 것이라고 작성하였다. 또한 피드백의 구체성을 높이기 위해 창의성의 구체적인 영역인 독창성, 유연성 등을 명시하였다. 인공지능 조건의 피드백은 다음과 같으며, 사람 조건에서는 '인공지능' 대신 '사람 전문가'로 수정하였고 다른 내용은 동일하게 설정하였다.

> "창의적 과제에 대한 귀하의 답변을 인공지능이 정밀히 분석하였습니다. 심사 기준은 클립을 사용한 과제의 경우, 독창성, 유연성, 유창성, 정교함을 토대로 심사하였고, 3가지의 물건을 이용해 장난감 만들기 과제의 경우 실용성과 독창성을 기준으로 심사하였습니다. 결과를 꼼꼼하게 읽어주시길 바랍니다. 귀하의 답변은 다른 사람의 답변에 비해 창의성이 부족합니다. 창의성의 영역인 독창성, 유창성, 유연성, 정교함 측면에서 모두 평균보다 낮은 하위그룹에 속합니다. 인공지능이 창의성에 대해 판단한 결과, 귀하의 답변에는 창의성을 지니고 있다고 판단 될 만한 근거가 거의 없었습니다. 창의성 과제를 수행 시, 답변을 작성할 때 어려움을 느꼈을 것으로 생각하고, 이로 인해 좋은 답변을 생각하지 못한 것으로 여겨집니다."

현실적 위협과 정체성 위협

현실적 위협은 Huang 등(2021)이 사용한 척도를, 정체성 위협은 Yogeeswaran과 Dasgupta (2014)의 연구에서 개발하고 Złotowski 등(2017)에서 사용한 척도를 번안해 사용하였다. 현실적 위협을 측정하는 두 문항, 정체성 위협을 측정하는 두 문항을 사용하였고, 리커트 식 5점 척도(1점 = 매우 그렇지 않다, 5점 = 매우 그렇다)로 측정하였다. 현실적 위협의 예시 문항은 "나의 기술로 할 수 있는 일이 인공지능(전문가)으로 대체될 까봐 걱정된다"가 있으며, 정체성 위협의 예시 문항으로는 "인공지능(전문가)은 인간의 고유성을 위협한다."가 있다. 본 연구에서 현실적 위협과 정체성 위협의

Spearman-Brown 상관 계수는 각각 .75와 .52였다.

지각된 원천의 신뢰성

지각된 원천의 신뢰성은 Sah 등(2013)의 척도를 번안하여 사용하였다. Sah 등(2013)은 지각된 원천의 신뢰성을 일관성 측면과 신뢰성 측면으로 나누어 측정하였는데, 본 연구에서는 피드백을 한 번만 제공하기 때문에 일관성과 관련된 문항은 제외하고 신뢰성 차원에 속한 총 다섯 개의 문항으로 측정하였다. 다섯 개의 문항은 리커트식 7점 척도(1점 = 매우 그렇지 않다, 7점 = 매우 그렇다)로 측정하였다. 예시 문항은 "나는 피드백을 제공한 인공지능(사람)을 신뢰한다"가 있다. 본 연구에서 해당 척도의 내적 일관성 신뢰도 계수는 .86이었다.

피드백 수용도

피드백 수용도는 Sah 등(2013)의 연구에서 사용한 다섯 개의 문항을 번안하여 측정하였다. 리커트식 7점 척도(1점 = 매우 그렇지 않다, 7점 = 매우 그렇다)로 측정하였으며 예시 문항으로는 "나는 피드백 결과를 수용한다." 및 "나는 피드백 내용에 대해 동의한다." 등이 있다. 본 연구에서 해당 척도의 내적 일관성 신뢰도 계수는 .67이었다.

내적 통제소재

내적 통제소재에 대한 척도는 Levenson(1981)의 연구에서 사용한 문항을 번안하여 총 여덟 개의 문항으로 측정하였으며, 6점 척도(1 점 = 매우 그렇지 않다, 6점 = 매우 그렇다)로 측정하였다. 예시 문항으로는 "나는 계획을 세울 때 계획한 일이 반드시 달성될 것이라고 거의 확신한다."가 있다. 본 연구에서 해당 척도의 내적 일관성 신뢰도는 .79이었다.

자기수행평가

첫 번째 설문에서 두 개의 창의적 과제를 수행한 후 참가자들로 하여금 자신의 수행에 대해서 평가를 하도록 하였다. 자신의 수행을 평가할 수 있는 두 개의 질문에 응답하도록 하였다. "두 가지의 창의적 과제에 대해서 본인 스스로 영점에서 100점 중 몇 점이라고 생각합니까?"에 대해 질문하고 영에서 100까지 정수로 표기할 수 있도록 하였다. 또 다른 질문은 "두 가지의 창의적 과제에 대해 본인 스스로 얼마나 잘 수행했다고 생각합니까?"이며, 리커트식 5점 척도(1점 = 매우 못함, 5점 = 매우 잘함)로 측정하였다.

분석 방법

본 연구의 자료 분석을 위해 SPSS 28.0 프로그램을 사용하였다. 가설을 검증하기에 앞서 두 번째 설문의 결측치가 무선으로 발생한 것을 확인하기 위해 첫 번째 설문에만 응답한 사람과 두 설문 모두 응답한 사람들의 인구통계학적 변인과 내적 통제소재 수준 및 자기수행 평가 수준을 독립표본 *t*-검증을 통해 분석하였다. 다음으로, 인공지능 기술과 사람이 제공한 부정적 피드백에 대한 수용도, 현실적 위협, 정체성 위협, 신뢰성에 미치는 주효과를 검증하기 위해 독립표본 *t*-검증을 실시하였다. 현실적 위협, 정체성 위협, 신뢰성의 매개 효과는 Hayes(2013)의 PROCESS Macro Model 4를 활용하여 다중 매개 분석을 진행하였다. 이후, 매개 효과의 유의성 검증을 위해 5,000회의 부트스트래핑(bootstrapping)을 실시하였다. 부트

스트래핑 검증 결과, 95% 신뢰구간 내에 영이 포함되지 않으면, 간접 효과가 존재한다고 판단하였다. 마지막으로, 내적 통제소재의 조절 효과를 확인하기 위해서 위계적 회귀 분석을 진행하였다.

결 과

가설 검증에 앞서 설문 2의 결측 자료가 무선적으로 발생했는지 확인하기 위해 첫 번째 설문에 응답한 사람들(n = 79)과 두 설문 모두 응답한 사람들(n = 230)의 연령, 교육 수준, 내적 통제소재, 창의적 과제의 수행에 대한 자기 평가 수준을 비교하였다. 그 결과, 첫 번째 설문에 응답한 집단의 나이, 교육 수준, 성별은 두 설문 모두 응답한 사람의 나이(p = .92), 교육 수준(p = .42), 성별과 유의한 차이를 보이지 않았다. 또, 내적 통제소재에서도 두 집단 간 차이는 유의하지 않았으며(p = .20), 창의적 과제의 수행 평가 수준 역시, 두 집단의 차이가 유의하지 않았다(ps < .42). 이를 통해, 두 번째 설문에서 발생한 결측은 무선적이라는 것을 알 수 있다.

인공지능과 사람의 부정적 피드백이 위협, 신뢰성, 수용도에 미치는 영향

가설 검증 전 피드백 수용도에 영향을 줄 수 있는 최종학력, 전공, 내적 통제소재 수준 및 자기 수행평가 수준이 인공지능 및 사람 피드백 조건 간 차이가 있는지 확인하고자 t-검증과 카이제곱 검증을 실시하였다. 인공지능 피드백 집단에는 130명, 사람 피드백 집단에는 100명이 포함되었으며, 인공지능 피드백 집단과(M = 2.75, SD = 0.76) 사람 피드백 집단의(M = 2.68, SD = 0.68) 최종 학력의 차이는 유의하지 않았다, t = 0.77, p = .45. 전공은 인문, 사회, 공학, 예체능, 기타로 나누어 측정하였으며, 기존 연구에서 공학 계열 전공자들은 인공지능 기술에 대해 긍정적인 태도와 수용적 태도를 보이기 때문에(Zhang & Dafoe, 2019), 공학 계열과 비공학 계열로 나누어 집단 간 차이를 살펴보았다. 인공지능 조건에서 공학 계열은 20.16%, 사람 조건에서 공학 계열은 22.00%로 실험 조건 간 해당 전공의 차이는 유의하지 않았다, $\chi^2(1)$ = .02, p = .89. 추가적으로 창의성과 관련된 일을 할 가능성이 높은 예체능 계열과 비예체능 계열의 차이도 살펴봤지만, 인공지능 조건(7.75%)과 사람 조건(8.00%) 간 차이가 유의하지 않았다, $\chi^2(1)$ = .001, p = .99. 내적 통제소재와 자기수행평가 수준 역시, 인공지능 피드백 집단과(M = 4.22, SD = 0.68) 사람 피드백 집단(M = 4.19, SD = 0.61)의 내적 통제소재 수준 역시 유의한 차이를 보이지 않았다, $t(228)$ = 0.28, p = .78. 수행 수준의 자기 평가에 있어서도, 100점으로 측정했을 때에 인공지능 집단과(M = 48.46, SD = 24.87) 사람 피드백 집단(M = 44.34, SD = 23.04) 간 차이는 유의하지 않았으며, $t(228)$ = 1.29, p = .23. 리커트 5점 척도로 측정했을 때도 인공지능 집단(M= 2.62, SD = 0.89)과 사람 피드백 집단(M = 2.61, SD = 0.82) 간 차이가 유의하지 않았다, $t(228)$ = 0.11, p = .46. 이를 통해 집단 간 인구통계학적 요인 및 피드백 수용도에 영향을 줄 수 내적 통제소재 및 자기 수행평가 점수는 무선적이라는 것을 알 수 있다.

각 변인 간 상관관계를 분석하여 표 1에 제시하였다. 피드백 원천은 현실적 위협(r =

-.30), 정체성 위협과(r = -.36) 부적 상관을 보였으며, 신뢰성과 상관 관계는 유의하지 않았다(r = .03). 피드백 수용도는 현실적 위협과 정적 상관을(r = .17), 정체성 위협과는 유의하지 않은 관계를(r = -.04), 신뢰성 인식과는 정적 상관을 보였다(r = .67).

피드백 원천에 따른 피드백 수용도, 위협, 신뢰성을 확인해 보았다. 인공지능 피드백 집단의 피드백 수용도와(M = 4.82, SD = 0.96) 사람 피드백 집단의 피드백 수용도의(M = 4.86, SD = 1.03) 차이는 유의하지 않았으며, 가설 1은 지지되지 않았다, t(228)= -0.30, p = .76, Cohen's d = -0.04. 예측한 대로 인공지능 피드백 집단의 현실적 위협 수준은(M = 3.43, SD = 1.04) 사람 피드백 집단의 현실적 위협 수준보다(M = 2.78, SD = 1.03) 유의하게 높았다, t(228) = 4.72, p < .001, Cohen's d = 0.63. 정체성 위협 역시 인공지능 기술 기반의 부정적 피드백을 받은 집단이(M = 3.38, SD = 0.90) 사람 피드백을 받은 집단보다(M = 2.69, SD = 0.90) 더 많이 느꼈다, t(228)= 5.76, p < .001, Cohen's d = 0.77. 따라서, 가설 2a와 가설 2b는 지지되었다. 반면, 인공지능 피드백 집단의 신뢰성 수준과(M = 3.18, SD = 0.81) 사람 피드백 집단의 신뢰성의(M = 3.23, SD = 0.77) 차이는 유의하지 않았다, t(228)= -0.46, p = .65, Cohen's d = -0.06. 따라서 가설 4는 지지되지 않았다. 실험 조건 간 신뢰성 차이가 유의하지 않기 때문에, 신뢰성의 매개 효과를 가정한 가설 5를 검증할 수 없고, 이에 가설 5는 지지되지 않았다.

추가로, 성별, 나이, 자기 수행 점수를 통제했을 때도 결과가 달라지는지 확인하였으나, 결과는 달라지지 않았다. 즉, 현실적 위협과, F(1, 224) = 22.70, p < .001, 정체성 위협은, F(1, 224) = 23.62, p < .001, 사람 피드백 대비 인공지능 피드백 조건에서 높았으나 신뢰성과, F(1, 224) = 0.01, p = .90, 피드백 수용

표 1. 기술 통계치 및 상관 관계

	M	SD	1	2	3	4	5	6	7	8	9
1. 나이	29.73	5.30	-								
2. 성별	-	-	-.09	-							
3. 피드백 원천	0.43	0.50	-.08	-.08	-						
4. 자기수행평가(1)	46.67	24.13	.05	-.03	-.09	-					
5. 자기수행평가(2)	2.62	0.86	-.01	.01	-.01	.82**	-				
6. 현실적 위협	3.15	1.08	.10	.02	-.30**	.14*	.11	-			
7. 정체성 위협	3.08	0.96	.04	.07	-.36**	.19**	.15*	.62**	-		
8. 신뢰성 인식	3.20	0.79	.03	-.19**	.03	-.15*	-.20**	.10	-.06	-	
9. 피드백 수용도	4.84	0.99	-.02	-.06	.02	-.24**	-.25**	.17**	-.04	.67**	-
10. 내적 통제소재	4.20	0.65	-.05	-.15*	-.02	.27**	.15*	.10	.05	.11	.12

주. N = 230. 피드백 원천은 인공지능 피드백을 0으로 사람 피드백을 1로 코딩함. 성별은 남성을 0으로 여성을 1로 코딩함. * p < .05, ** p < .01.

도는, $F(1, 224) = 0.003$, $p = .95$, 집단 간 유의한 차이가 없었다.

위협의 매개 효과

신뢰성은 실험 조건 간 유의한 차이를 보이지 않아, 현실적 위협과 정체성 위협의 매개 효과만 검증하였다. 실험 조건에 따른 피드백 수용도의 차이는 유의하지 않았지만, 독립 변인과 종속 변인의 관계가 유의하지 않아도 매개 효과를 검증할 수 있다는 연구자들의 의견에 따라(Hayes, 2009; Shrout & Bolger, 2002), 매개 효과를 분석하였다. 다중 매개 분석을 진행하였고, 통제 변인으로 성별, 나이, 자기 수행 평가 점수를 투입하고 다중 매개 효과 분석을 진행하였다. 그 결과(그림 2), 인공지능 피드백은 사람의 피드백에 비해 현실적 위협(β = -.64, $p < .001$) 및 정체성 위협을(β = -.66, $p < .001$) 높였다. 매개 변인과 종속 변인의 관계를 살펴본 결과, 정체성 위협을 많이 느낄수록 피드백 수용도는 낮아졌으나(β = -.16, $p = .05$), 현실적 위협 수준이 높을수록 피드백 수용도는 오히려 높아졌다는 것을 확인했다(β = .29, $p < .001$).

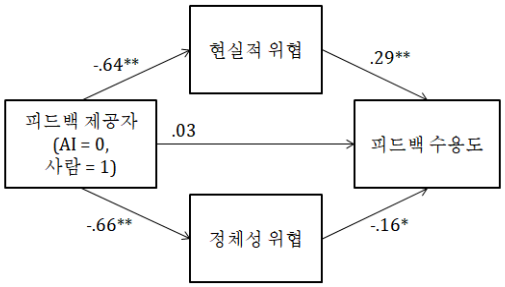


그림 2. 현실적 위협과 정체성 위협의 매개 효과. 주. 그림에 포함된 회귀 계수는 표준화된 회귀 계수임.

위협의 간접 효과를 파악하기 위해 부트스트래핑 분석을 실시하였다. 현실적 위협의 간접 효과는 신뢰 구간에 영을 포함하지 않아 간접 효과가 관찰되었다, $B = -0.19$, Boot SE = 0.07, 95% CI [0.34, -0.08]. 인공지능의 부정적 피드백이 인간의 부정적 피드백보다 현실적 위협을 증가시키고 이는 피드백 수용도를 높였다. 간접 효과는 밝혀졌지만 가설 방향이 예상과 달라 가설 3a는 지지되지 않았다. 정체성 위협의 간접 효과도 관찰되었다, $B = 0.11$, 95% CI [0.01, 0.26].

인공지능 피드백은 사람 피드백에 비해서 정체성 위협 수준이 높았고, 정체성 위협이 피드백 수용도를 낮추는 원인이 되었으며 이는 가설 3b를 지지하는 결과이다. 실험 조건이 피드백 수용도에 미치는 직접 효과는 확인되지 않았으며, $B = 0.08$, Boot SE = 0.14, 95% CI [-0.20, 0.34], 현실적 위협과 정체성 위협의 간접 효과를 감안했을 때 완전 매개하는 것을 알 수 있다. 나이($p = .46$), 성별($p = .35$), 자기 수행 평가($ps. < .17$)는 피드백 수용도에 유의한 영향을 미치지 않았다.

추가적으로 가설로 설정하지는 않았지만, 신뢰성과 피드백 수용도의 정적 관계를 기반으로 정체성, 현실성 위협을 거쳐 신뢰성을 통해 피드백 수용도로 이어지는 경로를 구조방정식 모형을 통해 분석을 진행해보았다. AMOS 21 프로그램을 통해 분석하였으며, 실험 조건이 정체성, 현실성 위협을 통해서 신뢰성에 영향을 주고, 신뢰성을 거쳐서 피드백 수용도에 영향을 주는 경로와 간접 효과의 유의성을 확인해 보았다. 그 결과, 인공지능 피드백 조건에서 사람 피드백 조건보다 현실성 위협(β = -.33, $p < .001$), 정체성 위협(β = -.41, $p < .001$)을 많이 느꼈다. 정체성 위협은

신뢰성(β = -.20, p = .003)에 부적 영향을, 현실적 위협은 신뢰성에 정적 영향을(β = .24, p = .001) 미쳤으며 신뢰성은 피드백 수용도에 정적 영향을 미쳤다(β = .64, p < .001). 현실성 위협과, 95% CI [0.26, -0.04], 정체성 위협이, 95% CI [0.04, 0.29]. 신뢰성을 통해서 피드백 수용도에 영향을 주는 간접 효과도 각각 관찰되었다. 이를 통해 현실적 위협과 정체성 위협이 피드백 수용도에 미치는 각기 다른 효과가 본 분석에도 확인되었다. 가설로 설정하지는 않았지만 위협이 신뢰성에 영향을 주고 이것이 피드백 수용도에 영향을 줄 수 있음을 알 수 있다.

내적 통제소재의 조절 효과

마지막으로 내적 통제소재의 조절 효과를 확인하였다. 본 연구에서는 인공지능 기술이 제공하는 부정적 피드백으로 인한 위협과 피드백 수용도의 관계에서 내적 통제소재의 조절 효과에 초점을 두었고 조절 효과 분석은 인공지능 피드백 조건의 자료를 대상으로 진행하였다. Aiken과 West(1991)의 방법에 따라, 위계적 회귀분석을 실시하였으며, 1단계에서는 통제 변인인 나이, 성별, 자기 수행 평가 점수를, 2단계에서는 현실적 위협, 정체성 위협, 내적 통제소재, 3단계에서는 현실적 위협 및 정체성 위협 수준과 내적 통제소재의 상호작용항을 투입하였다. 다중공선성의 영향을 최소화하기 위해 위협 수준 및 내적 통제소재는 평균 중심화하였다. 분석 결과, 현실적 위협과 내적 통제소재의 상호작용이 피드백 수용도에 미치는 영향은 유의하지 않았다(β = -.07, p = .50). 이에, 가설 6a는 지지되지 않았다. 반면, 정체성 위협과 피드백 수용도의 관계에서 내적 통제소재의 조절 효과는 유의하였으며(β = -.27, p < .02), 정체성 위협과 내적 통제소재의 상호작용항은 피드백 수용도의 3.71%를 설명하는 것으로 나타났다. 즉, 가설 6b는 지지되었다.

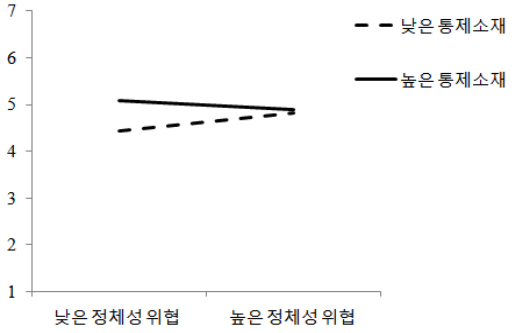


그림 3. 정체성 위협과 내적 통제소재의 상호작용이 인공지능 피드백 수용도에 미치는 영향

구체적으로 상호 작용의 양상을 확인하기 위해 단순 기울기 검증(simple slope analyses)을 실시하였다(그림 3). 내적 통제소재 수준이 높을 때는(M+1SD) 정체성 위협 수준이 피드백 수용도에 미치는 영향은 유의하지 않았다(b = -0.13, SE = 0.11, t = -1.17, p = .27). 반면, 내적 통제소재 수준이 낮을 때는(M-1SD), 정체성 위협이 올라갈수록 피드백 수용도가 높아졌다(b = 0.25, SE = 0.14, t = 1.87, p = .06). 내적 통제소재가 정체성 위협과 피드백 수용도의 관계에서 완화 역할을 하는 것을 알 수 있다. 구체적으로, 내적 통제소재 수준이 낮을 때 정체성 위협을 높게 느낄수록 피드백 수용도가 높아지는 정적 관계가 나타났다.

논 의

본 연구를 통해 국내 성인을 대상으로 창의

적 수행 과제에서 인공지능 기반의 부정적 피드백이 위협, 신뢰성 인식 및 피드백 수용도에 미치는 영향을 살펴보았다. 연구 결과, 인공지능 기반의 부정적 피드백을 받을 경우 사람의 피드백에 비해 현실적 위협, 정체성 위협이 높아졌으며, 정체성 위협은 인공지능의 피드백 수용도를 낮추는 요인이 되었다. 반면, 인공지능 기반의 부정적 피드백을 받은 경우 현실적 위협 수준은 높아졌지만, 현실적 위협이 높아질수록 인공지능의 부정적 피드백의 수용도가 올라갔다. 조절 효과 검증 결과, 내적 통제소재 수준은 정체성 위협과 피드백 수용도의 관계를 조절하는 것으로 나타났다. 가설의 방향과 달랐지만 내적 통제소재 수준이 낮을 때만 정체성 위협이 피드백 수용도를 높였으며, 내적 통제소재 수준이 높을 때는 정체성 위협과 피드백 수용도의 관계가 유의하지 않았다.

본 연구의 의의는 다음과 같다. 첫째, 동일한 피드백이라고 해도 사람이 제공할 때보다 인공지능이 부정적인 피드백을 제공할 경우 더 현실적, 정체성 위협을 느낀다는 것을 확인하였다. 창의적 수행 과제에 대한 인공지능의 부정적 피드백은 현실적 위협과 정체성 위협을 느끼게 하는 요인이 될 수 있음을 시사한다. 이 결과는 인공지능의 판단이나 인공지능으로 변화할 사회를 생각했을 때 현실적, 정체성 위협이 높아진다는 기존 연구 결과와 일치한다(Huang et al., 2021; Jackson et al., 2020; Park et al., 2024). 인공지능 기술의 판단에 대한 사람들의 정서적 반응은 다르며 위협은 대표적인 정서적 반응인 것으로 알려져 있다. 특히, 본 연구에서는 직무 대체를 직접적으로 언급하지 않은 상황이었음에도 인공지능 기술이 적용된 피드백을 읽을 때 현실적 위협을 느낀다는 결과를 통해서 현실적 위협은 인공지능 기술과 관련된 결과물을 대할 때 사람들이 보편적으로 보이는 정서적 반응임을 알 수 있다. 기존 연구에서도 인공지능 및 로봇의 적용 수준의 국가 간 비교를 통해서 자신의 직종이 직접적으로 대체되지 않더라도 사회에서 로봇 적용 수준이 높아지면 현실적 위협 수준이 높아졌다(Yam et al., 2023). 자신이 속한 직종이나 조직이 실제로 대체되거나 없어진다는 객관적인 사실보다, 사회가 인공지능 기술에 의해 변화하고 있다는 주관적인 판단이 현실적 위협을 높인다는 주장과 일치하는 결과이다(Shoss, 2017; Yam et al., 2023).

둘째, 인공지능과 사람의 부정적 피드백에 대해서 피드백 수용도와 원천의 신뢰성 수준은 유의한 차이가 나지 않았다는 것을 확인했다. 이는 서구권에서 진행된 기존 연구와 일치하지 않는 결과이며(예: Dietvorst et al., 2015; Jussupow et al., 2022), 부분적으로 문화적 차이에 기인되었을 것으로 추측된다. 동아시아권에서는 인공지능 기술을 의인화하는 경향이 서구권보다 높으며(Folk et al., 2023), 미국이나 북유럽 국가보다 일터에서의 인공지능 사용에 대해서 덜 부정적으로 평가하는 경향이 있다(Mantello et al., 2023). 한국인을 대상으로 진행한 연구에 따르면, 인공지능과 사람을 비교했을 때 인공지능에 대한 부정적인 반응을 넘어서 인공지능의 판단을 사람의 판단보다 더 신뢰하는 결과도 나타났다. 예를 들어, 인공지능 의사에 대한 신뢰 수준이 사람 의사에 대한 신뢰보다 높았고(Kim & Kim, 2022) 채용 지원자를 대상으로 한 연구에서도 인간 채용담당자보다 인공지능 채용을 더 공정하다고 믿는 경향이 있었다(민지현 등, 2018). 인공지능 기

술에 대한 의인화 경향 및 문화를 통해 습득된 믿음과 경험이 인공지능 기술 기반의 피드백에 대한 신뢰성 및 피드백 수용도에 영향을 준 것이 아닌가 한다. 정서적으로 위협을 느끼더라도 유용성과 같은 인지적인 평가 등이 원천에 대한 신뢰성에 영향을 주는 것으로 추정한다.

셋째, 인공지능에 대한 정체성 위협과 현실적 위협이 피드백 수용도에 차별적인 영향을 준다는 것을 확인하였다. 정체성 위협을 느낄 경우 해당 피드백을 수용하지 않게 되며 이는 정체성 위협이 인공지능에 대한 부정적인 태도를 예측한다는 기존 연구들과 일치하는 결과이다(Huang et al., 2021; Jackson et al., 2019). 하지만, 흥미롭게도 현실적 위협은 인공지능의 부정적 피드백의 수용도를 높이는 원인이 되었다. 일반적으로 위협을 느낄 경우 회피 경향성이 촉발되며 대상이나 사건에 대한 부정적인 태도로 이어지지만, 본 연구에서는 현실적 위협을 느낄 경우 인공지능 기반의 피드백을 더 수용하는 것으로 나타났다. 이는 가설과는 반대되는 결과이다. 일반적으로 위협을 느꼈을 때 회피하거나(flight) 싸울 수도(fight) 있지만, 위협은 오히려 방어적인 노력(protective effort)을 통해 위협을 주는 대상에 주의를 쏟고 건설적으로 행동하게 만드는 요인이 되기도 한다(Lebel, 2017). 현실적 위협을 느낄 경우 수용하지 않는 방향으로 피하거나 싸우기도 하지만 어떤 경우에는 방어적인 노력을 통해 위협이 해당 위협이 초래하는 변화에 대해 관심을 갖고 그 변화를 적극적으로 활용하는 선제적 행동을 향상시킬 수도 있다(Lebel, 2017). 특히, 경제적 불안정성이 동반하는 위협은 위협에 대처하기 위해 새로운 기술을 쌓는 것과 같은 선제적 행동을 촉발시키기도 했다(Lebel et al., 2023). 위협이라는 일반적인 특성은 위협을 제공하는 대상이나 사건으로부터 멀어지게 하지만 현실적 위협은 경제적 어려움에 대한 생각을 동반하고 이는 피하는 것이 아니라 수용함으로써 대처해야 한다는 태도를 불러일으킨 것으로 파악된다.

또, 두 종류의 위협의 피드백 수용도에 상반된 영향을 미치는 것은 두 위협의 대처방식이 다를 수 있음을 시사한다. 이는 Davis(2012)가 주장한 기술에 대한 사람의 양가적인 태도를 반영하는 것일 수 있다. Davis(2012)는 사람들은 기술의 유용성을 긍정적으로 평가하지만 인간 관계의 대체 가능성 등에 대해서는 부정적으로 평가하며 기술에 대해서 긍정과 부정의 양가적인 태도를 보인다고 설명한다. 창의적 수행에 대한 인공지능 기반의 피드백에 대해서 현실적 위협을 느낄 정도로 질 좋은 결과를 도출한다고 느낄 경우 유용성을 느끼고 오히려 피드백을 수용함으로써 일터에 적응하고자 하는 태도를 보이는 것일 수도 있다. 하지만 인간다움을 위협하는 정체성 위협을 느낄 경우는 수용도에 긍정적인 영향을 미치지는 못하는 것으로 보인다.

본 연구 결과의 특징적인 점 중 하나는 피드백 원천이 두 종류의 피드백에 미치는 영향이 비슷하게 나타났다는 것이다. 회귀 계수 등에 있어서 피드백 원천이 두 종류의 위협에 미치는 영향이 비슷했고 오히려 Cohen's *d*로 확인된 효과 크기에 있어서 정체성 위협의 집단 간 차이가 현실적 위협의 집단 간 차이보다 다소 큰 것으로 나타났다. 기존 연구에서 두 종류의 위협을 함께 살펴본 경우 일반적으로 현실적 위협의 영향력이 다소 큰 것으로 나타났고, 두 위협이 자율 로봇에 대한 부정적인 태도에 미치는 이중 매개 효과를 살펴본

경우 현실적 위협만 유의한 것으로 나타나기도 했다(Złotowski et al., 2017). 이에, 연구자들은 자율 로봇에 대해 갖는 부정적인 태도의 주요 원인은 현실적 위협 때문일 것이라고 설명한 바 있다(Złotowski et al., 2017). 이런 주장과 달리, 본 연구에서 상대적으로 정체성 위협 수준이 높은 것은 창의적인 과업이라는 특성 때문일 수 있으며, 이는 과업의 특성에 따라 인공지능 기술에 대한 세부적인 위협의 수준이 다를 수 있음을 뜻한다. 또, 위협의 종류가 피드백 수용도에 미치는 영향력이 다르다는 것을 감안할 때 인공지능 기술에 대한 위협의 종류를 세분화하여 이해하는 것이 필요함을 시사한다.

이러한 연구 결과들을 통해 창의적 과제 수행 시 인공지능 기반의 피드백을 받을 경우 사람이 제공하는 피드백과 다른 인지적, 정서적 반응을 보인다는 것을 알 수 있다. 이런 연구 결과는 인공지능 기술이 피드백 원천으로서 작동할 수 있음을 보여주며, 이를 통해 기존 피드백 연구 결과를 확장할 수 있다. 피드백 원천은 피드백 수용도에 있어서 주요한 요인인데 피드백 원천에 대한 신뢰성, 전문성에 대한 인지적인 요인이 원천의 특징을 설명하는 주요 요인으로 간주되어 왔다(Ilgen et al., 1979; Reinke, 2003). 몇몇 연구에서 원천의 지위(status)나 원천에 대한 관계의 질을 연구했지만(Lechermeier & Fassnacht, 2018), 전문성, 신뢰 같은 인지적 요인보다 사회적, 정서적 요인에 대한 연구는 상대적으로 관심이 적은 분야였다. 특히, 원천에 대한 위협과 같은 부정적인 정서에 대한 피드백 연구는 드문 상황이다. 본 연구 결과 및 기존 연구 결과를 통해 인공지능 기반 피드백에 있어서는 정서적 반응 역시 중요하다는 것을 알 수 있다. 이런 결과는 인공지능 기반 피드백 및 피드백 수용도를 이해하기 위해서는 피드백 원천에 대한 인지적, 정서적, 사회적 측면과 같은 다각적인 측면을 고려해야 함을 시사한다.

마지막으로, 본 연구에서 내적 통제소재가 인공지능 기반 피드백의 수용도에 영향을 줄 수 있음을 확인하였다. 가설에서는 내적 통제소재가 정체성 위협과 피드백 수용도의 부적 관계를 약화시킬 것이라고 가정하였지만, 본 연구에서는 내적 통제소재가 낮은 사람에게 정체성 위협과 피드백 수용도의 정적 관계가 높아지는 것으로 나타났다. 위협을 느꼈을 때 통제감이 없고 불확실성이 큰 상황일수록 위협의 영향을 더 크게 받는다는 것을 감안하면(Lerner & Keltner, 2001), 내적 통제소재가 낮은 사람이 정체성 위협을 느끼면 통제감을 잃고 불확실성을 더 크게 경험할 수 있다. 불확실성으로 인해서 피드백을 덜 수용할 것이라고 예상했으나 오히려 위협이 피드백을 수용하게 하는 기제가 되는 것을 알 수 있다. 자신의 통제감이 흔들릴 경우 피드백을 거부하는 것이 아니라 다소 무비판적으로 수용하면서 위협에 대응하는 것으로 나타났다. 이는 변화에 저항을 하는 것은 때로는 자원을 필요로 하는 일로서 현재 위협을 느끼고 통제감이 없을 경우 피드백을 수용하지 않는 것보다 수용하는 것이 더 손쉬운 대처이자, 인상 관리의 차원에서도 더 나은 대처 방안으로 간주되기 때문에(Miller & Mulligan, 2002) 피드백 수용도가 높아지는 것으로 추정한다.

또, 내적 통제수준이 높을수록 위협과 피드백 수용도의 정적 관계가 약할 것이라고 생각했지만, 내적 통제소재 수준이 높은 사람들에게는 위협과 피드백 수용도의 관계는 유의하지 않았다. 가설에서는 내적 통제소재가 높은

사람들은 새로운 것을 적극적으로 학습한다는 특성 때문에 인공지능 기반 피드백 수용도가 높을 것이라고 예상했지만, 내적 통제수준이 높은 사람들은 위협과 불안이 직접적으로 피드백 수용도를 높이기보다 위협을 진단하는 등의 과정이 개입되는 것으로 파악된다. 내적 통제소재가 높은 사람들은 상황이 본인의 통제 하에 있다고 믿으며, 위협 상황에서도 신중하게 행동을 보이는 것으로 나타났다(Miller & Mulligan, 2002). 내적 통제소재가 위험 추구 행동에 미치는 영향에 대한 연구에 따르면, 위협을 느낀 경우 내적 통제수준이 낮은 사람들은 위험 추구 행동을 많이 하지만, 내적 통제수준이 높은 사람들은 위험을 진단하고 위험 추구 행동은 덜 하는 것으로 나타났다(Miller & Mulligan, 2002). 즉, 내적 통제소재가 높은 사람들은 위협에 즉각적인 영향을 받기보다는 위협의 상황을 파악하고 판단을 기반으로 행동하는 경향이 있고 이러한 특성이 본 연구 결과에 반영된 것으로 추정된다.

본 연구 결과를 기반으로 한 실용적 함의는 다음과 같다. 첫째, 창의적 과제 수행에 있어서 인공지능 기반의 피드백을 사용할 경우, 인공지능의 피드백이 야기할 수 있는 현실적 위협과 정체성 위협을 줄일 방안을 모색하여야 한다. 정체성 위협이 높아질 경우 피드백 수용도가 낮아진다는 것을 감안하여 정체성 위협의 부정적인 영향을 완화시키는 방안이 필요하다. 인공지능 기술에 대한 사용 경험이 많아질수록 긍정적인 태도를 보이고 위협을 덜 느낀다는 결과에 기반해(Zhang & Dafoe, 2019), 인공지능 기술 활용 방법 그리고 인공지능의 역할과 한계를 알릴 필요가 있을 것이다. 또, 현실적 위협은 피드백 수용도를 높이지만 현실적 위협은 정서 소진 및 스트레스를 유발할 수 있으며, 위협 자체보다는 위협이 동반하는 예방적 노력이 선제적 행동을 일으키기 때문에(Lebel, 2017), 조직에서는 현실적 위협을 관리할 필요가 있을 것이다. 현실적 위협은 직무에 대한 불안으로 인한 것이기 때문에 인공지능 기술을 조직에 도입할 때 조직 구성원을 의사 결정에 참여시키고, 조직 내 충분한 소통을 통해서 조직구성원이 통제감을 느끼게 하는 것이 필요할 것이다(Shoss, 2017).

또, 개인의 내적 통제소재의 영향력을 고려하여 차별화된 교육 및 개입을 할 필요가 있다. 내적 통제소재가 낮은 사람들은 정체성 위협에도 불구하고 인공지능의 피드백을 수용한다는 결과를 바탕으로 피드백 수용의 이유를 확인할 필요가 있다. 이들이 인공지능 기반의 피드백을 점검하고 비판적 사고를 통해 무분별하게 수용하지 않도록 교육하고 개입할 필요가 있다. 또, 인공지능 기술의 활용은 창의적 수행을 향상시킨다는 연구 결과를 바탕으로(Jia et al., 2024), 내적 통제소재가 높은 사람들에게는 인공지능 피드백을 창의적 과업을 수행할 때 사용하는 이점을 알릴 필요가 있을 것이다. 인공지능 기반 피드백을 조직 및 일터에서 적용할 때 내적 통제소재와 같이 개인 특성에 따라 교육 및 개입을 차별화하는 것이 필요할 수 있다.

제한점 및 향후 연구 방안

본 연구의 제한점 및 향후 연구를 위한 제언은 다음과 같다. 본 연구는 한국인 20, 30대를 대상으로 진행되었으며, 다양한 문화적 요인과 개인차 요인을 고려해도 본 연구 결과가 반복 검증되는지 확인할 필요가 있다. 동아시아 국가들은 유럽과 북미 국가들에 비해 인공

지능 기술에 대해서 긍정적인 태도를 보이며 (Folk et al., 2023; Mantello et al., 2023), 문화에 따라 인공지능 기술에 대한 태도 뿐 아니라 자기 개념에 있어서 차이가 있다(Brockner et al., 1996). 또, 직무 연한 혹은 기술 수준에 따라 인공지능 피드백 수용도가 다르며(Jia et al., 2024; Tong et al., 2021), 창의적 자기효능감 같은 개인의 성격과 창의적 능력과 같은 요인이 창의적 수행에 영향을 줄 수 있다(Choi, 2004). 이러한 문화적, 개인차 요인을 감안해도 본 연구 결과가 나타나는지 추후 연구에서 확인할 필요가 있을 것이다.

또 다른 제한점은 본 연구에서는 창의적 수행에 대한 부정적 피드백에만 초점을 두었다는 것이다. 긍정적 피드백을 제공하였을 때도 본 연구 결과가 나타난다면 피드백 유인가에 관계없이 피드백 원천에 따른 결과를 확인할 수 있을 것이다. 또, 창의적 수행에 대한 평가에 대한 피드백을 구체적으로 제공하고자 하였지만 참가자들이 창의성 수준을 낮고 높음이 아니라 창의성이 없다고 평가를 해석했을 수 있다. 이에 자신에게 특정 자질 혹은 능력이 없다고 인공지능이 판단한 것이 위협 및 수용도에 영향을 주는 요인이 되었을 수 있다. 피드백 조건을 창의성 없음, 낮은 창의성, 높은 창의성 평가 등의 조건으로 나누어 구체적으로 피드백에 대한 반응이 달라지는지 확인할 필요가 있을 것이다.

마지막 제한점은 정체성 위협의 측정 도구 및 개념에 대한 것이다. 본 연구에서는 과거 연구에서 타당화되고 인공지능 연구에서 사용한 위협을 측정하는 도구를 사용했지만, 위협을 각각 두 문항으로 측정하였으며 정체성 위협의 경우 Spearman-Brown 상관 계수가 상대적으로 낮았다. 추후 연구에서는 한국 직장인을 대상으로 정체성 위협에 대한 개념을 구체화하고 더 많은 문항으로 측정할 필요가 있을 것이다.

이와 같은 한계점에도 불구하고 본 연구는 국내 성인을 대상으로 인공지능 기술이 창의적 과제에 대한 피드백 원천으로 역할을 하며 인공지능에 대한 위협이 피드백 수용도에 영향을 미친다는 것을 확인했다는데 의의가 있다. 동일한 피드백이라고 해도 인공지능이 제공할 경우 사람과 대비하여 더 많은 정체성, 현실적 위협을 느끼며, 이는 피드백 수용도에 차별화된 영향을 미치는 것을 확인하였다. 추가적으로 내적 통제소재가 정체성 위협과 피드백 수용도의 관계를 조절하여 개인의 성격이 창의적 과제의 인공지능 기술에 대한 피드백에 영향을 미친다는 것을 발견하였다. 본 연구의 결과는 기존 서구권에서 진행된 인공지능 기술의 피드백에 대한 태도와 차별화되며, 인공지능 기술에 태도는 다면적이라는 것을 보여준다. 인공지능 기술이 적용되는 작업 환경에서 국내 성인의 태도와 반응을 이해하기 위한 지속적인 연구와 관심이 필요함을 알 수 있다.

참고문헌

김시내, 박지영 (2024). 인공지능 기술 적용에 대한 한국 직장인의 태도: 척도 개발 및 타당화 연구. 한국심리학회: 산업 및 조직, 37(2), 385-413. https://doi.org/10.24230/kjiop.v37i2.81-117

노동균 (2019년 4월 12일). 인간과 AI가 공동 제작한 의자는 어떤 모습? IT조선, https://it.chosun.com/news/articleView.html?idxn

o=2019041201979

도은영, 이국희, 정지은 (2021). 사건 관여도와 심각성이 인공지능 판결에 대한 수용도에 미치는 효과. 인지과학, 32(4), 16. http://dx.doi.org/10.19066/cogsci.2021.32.4.001

문혜진, 남상희 (2022). 인공지능을 활용한 비동시적 면접 연구의 현황과 과제. 한국심리학회: 산업 및 조직, 35(3), 385-413. https://doi.org/10.24230/kjiop.v35i3.385-413

민지현, 김시내, 박용욱, 손영우 (2018). A Comparative Study of Potential Job Candidates' Perceptions of an AI Recruiter and a Human Recruiter. 한국융합학회논문지, 9(5), 191-202. https://doi.org/10.15207/JKCS.2018.9.5.191

박지영, 정예슬 (2020). 인공지능의 유능함에 대한 인식이 직장인의 직무불안정성과 냉소주의에 미치는 영향: 일의 의미의 조절된 매개효과. 한국심리학회지: 산업 및 조직, 33(2), 145-175. https://doi.org/10.24230/kjiop.v33i2.145-175

Abrams, D., & Hogg, M. A. (1988). Comments on the motivational status of self esteem in social identity and intergroup discrimination. *European Journal of Social Psychology*, *18*(4), 317-334. https://doi.org/10.1002/ejsp.2420180403

Albright, M. D., & Levy, P. E. (1995). The effects of source credibility and performance rating discrepancy on reactions to multiple raters. *Journal of Applied Social Psychology, 25*(7), 577-600. https://doi.org/10.1111/j.1559-1816.1995.tb01600.x

Anseel, F., & Lievens, F. (2009). The mediating role of feedback acceptance in the relationship between feedback and attitudinal and performance outcomes. *International Journal of Selection and Assessment*, *17*(4), 362-376. https://doi.org/10.1111/j.1468-2389.2009.00479.x

Bannister, B. D. (1986). Performance outcome feedback and attributional feedback: Interactive effects on recipient responses. *Journal of Applied Psychology, 71*(2), 203-210. https://doi.org/10.1037/0021-9010.71.2.203

Bedemariam, R., & Wessel, J. L. (2023). The roles of outcome and race on applicant reactions to AI systems. *Computers in Human Behavior, 148,* 107869. https://doi.org/10.1016/j.chb.2023.107869

Brett, J. F., & Atwater, L. E. (2001). 360° feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied psychology, 86*(5), 930. https://doi.org/10.1037/0021-9010.86.5.930

Brockner, J., & Chen, Y. R. (1996). The moderating roles of self-esteem and self-construal in reaction to a threat to the self: evidence from the People's Republic of China and the United States. *Journal of Personality and Social Psychology, 71*(3), 603-615. https://doi.org/10.1037/0022-3514.71.3.603

Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts, 12*(2), 177-192. https://doi.org/10.1037/aca0000136

Chiu, F.-C. (2012). Fit between future thinking and future orientation on creative imagination.

Thinking Skills and Creativity, 7(3), 234-244. https://doi.org/10.1016/j.tsc.2012.05.002

Choi, J. N. (2004). Individual and contextual predictors of creative performance: The mediating role of psychological processes. *Creativity Research Journal, 16*(2-3), 187-199. https://doi.org/10.1080/10400419.2004.9651452

Crocker, J., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review,* 108(3), 593. https://doi.org/10.1037/0033-295x.108.3.593

Cohn, G. (2018, October 25). AI Art at Christie's Sells for $432,500, *The New York Times.* https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html

Davis, J. L. (2012). Social media and experiential ambivalence. *Future Internet, 4*(4), 955-970. https://doi.org/10.3390/fi4040955

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114-126. https://doi.org/10.2139/ssrn.2466040

du Sautoy, M. (2019). Can AI ever be truly creative? *New Scientist, 242*(3229), 38-41. https://doi.org/10.1016/S0262-4079(19)30840-1

Feather, N. T. (1968). Change in confidence following success or failure as a predictor of subsequent performance. *Journal of Personality and Social Psychology, 9*(1), 38-46. https://doi.org/10.1037/h0025671

Fodor, E. M., & Carver, R. A. (2000). Achievement and power motives, performance feedback, and creativity. *Journal of Research in Personality, 34*(4), 380-396. https://doi.org/10.1006/jrpe.2000.2289

Folk, D., Wu, C., & Heine, S. (2023). *Cultural variation in attitudes towards social chatbots*. PsyArXiv. https://doi.org/10.31234/osf.io/wc895

Gervás, P. (2019). Exploring quantitative evaluations of the creativity of automatic poets. In F. Veale, T., Cardoso (Ed.), Computational Creativity. Computational synthesis and creative systems (pp. 275-304). Springer, Cham. https://doi.org/10.1007/978-3-319-43610-4_13

Guilford, J. P. (1959). Three faces of intellect. *American Psychologist, 14*(8), 469-479. https://doi.org/10.1037/h0046827

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*(4), 408-420. https://doi.org/10.1080/03637750903310360

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis.* New York, NY: The Guilford Press.

Heine, S. J., & Lehman, D. R. (1997). The cultural construction of self-enhancement: an examination of group-serving biases. *Journal of Personality and Social Psychology, 72*(6), 1268-1283.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology, 53*(1), 575-604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Higgins, E. T. (1987). Self-discrepancy: a theory relating self and affect. *Psychological Review, 94*(3), 319-340.

Huang, H. L., Cheng, L. K., Sun, P. C., & Chou,

S. J. (2021). The effects of perceived identity threat and realistic threat on the negative attitudes and usage intentions toward hotel service robots: the moderating effect of the robot's anthropomorphism. *International Journal of Social Robotics, 13,* 1599-1611. https://doi.org/10.1016/j.techsoc.2020.101410

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*(4), 349-371. https://doi.org/10.1037/0021-9010.64.4.349

Jackson, J. C., Castelo, N., & Gray, K. (2020). Could a rising robot workforce make humans less prejudiced?. *American Psychologist, 75*(7), 969-982. https://doi.org/10.1037/amp0000582

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. *Academy of Management Journal, 67*(1), 5-32. https://doi.org/10.5465/amj.2022.0426

Jussupow, E., Spohrer, K., & Heinzl, A. (2022). Identity threats as a reason for resistance to artificial intelligence: survey study with medical students and professionals. *JMIR Formative Research, 6*(3), e28750. https://doi.org/10.2196/28750

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15-25. https://doi.org/10.1016/j.bushor.2018.08.004

Kim, T. W., & Duhachek, A. (2020). Artificial intelligence and persuasion: A construal-level account. *Psychological Science,* 31(4), 363-380. https://doi.org/10.1177/0956797620904985

Kim, D. K. D., & Kim, S. (2022). What if you have a humanoid AI robot doctor?: An investigation of public trust in South Korea. *Journal of Communication in Healthcare, 15*(4), 276-285. https://doi.org/10.1080/17538068.2021.1994825

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin, 119*(2), 254. https://doi.org/10.1037/0033-2909.119.2.254

Landers, R. N., & Marin, S. (2021). Theory and technology in organizational psychology: A review of technology integration paradigms and their effects on the validity of theory. *Annual Review of Organizational Psychology and Organizational Behavior, 8*(1), 235-258. https://doi.org/10.1146/annurev-orgpsych-012420-060843

Lechermeier, J., & Fassnacht, M. (2018). How do performance feedback characteristics influence recipients' reactions? A state-of-the-art review on feedback source, timing, and valence effects. *Management Review Quarterly, 68(*2), 145-193. https://doi.org/10.1007/s11301-018-0136-8

Lebel, R. D. (2017). Moving beyond fight and flight: A contingent model of how the emotional regulation of anger and fear sparks proactivity. *Academy of Management Review, 42*(2), 190-206. https://doi.org/10.5465/amr.2014.0368

Lebel, R. D., Yang, X., Parker, S. K., & Kamran-Morley, D. (2023). What makes you proactive can burn you out: The downside of proactive skill building motivated by financial precarity and fear. *Journal of Applied Psychology, 108*(7), 1207-1222. https://doi.org/10.1037/apl0001063

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81*(1), 146-159. https://doi.org/10.1037/0022-3514.81.1.146

Levenson, H. (1981). Differentiating among internality, powerful others, and chance. *Research with the Locus of Control Construct, 1,* 15-63. https://doi.org/10.1016/b978-0-12-443201-7.50006-3

Longoni, C., & Cian, L. (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The "word-of-machine" effect. *Journal of Marketing, 86*(1), 91-108. https://doi.org/10.1177/0022242920957347

Magni, F., Park, J., & Chao, M. M. (2024). Humans as creativity gatekeepers: Are we biased against AI creativity? *Journal of Business and Psychology, 39,* 643-656. https://doi.org/10.1007/s10869-023-09910-x

Mantello, P., Ho, M. T., Nguyen, M. H., & Vuong, Q. H. (2023). Bosses without a heart: socio-demographic and cross-cultural determinants of attitude toward Emotional AI in the workplace. *AI & Society, 38*(1), 97-119.

Miller, R. L., & Mulligan, R. D. (2002). Terror management: The effects of mortality salience and locus of control on risk-taking behaviors. *Personality and Individual differences, 33*(7), 1203-1214. https://doi.org/10.1016/S0191-8869(02)00009-0

Nargesi, M. N. (2021). *Comparing AI and Human Feedback: The Role of Source in Feedback Perceptions* [Unpublished Doctoral dissertation] San Francisco State University. https://scholarworks.calstate.edu/downloads/r494vr478

Ng, T. W., Sorensen, K. L., & Eby, L. T. (2006). Locus of control at work: a meta analysis. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *27*(8), 1057-1087. https://doi.org/10.1002/job.416

Ng, A. (2018). *AI is the New Electricity.* O'Reilly Media.

Park, J., Woo, S. E., & Kim, J. (2024). Attitudes towards artificial intelligence at work: Scale development and validation. *Journal of Occupational and Organizational Psychology.* https://doi.org/10.1111/joop.12502

Reinke, S. J. (2003). Does the form really matter? Leadership, trust, and acceptance of the performance appraisal process. *Review of Public Personnel Administration, 23*(1), 23-37. https://doi.org/10.1177/0734371X02250109

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review, 10*(4), 336-353. https://doi.org/10.1207/s15327957pspr1004_4

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of

reinforcement. *Psychological monographs: General and Applied, 80*(1), 1-28. https://doi.org/10.1037/h0092976

Sah, S., Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes, 121*(2), 246-255. https://doi.org/10.1016/j.obhdp.2013.02.001

Scharowski, N., Perrig, S. A., Svab, M., Opwis, K., & Brühlmann, F. (2023). Exploring the effects of human-centered AI explanations on trust and reliance. *Frontiers in Computer Science, 5,* https://doi.org/10.3389/fcomp.2023.1151150

Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self‐Affirmation and the reduction of defensive biases. *Current Directions in Psychological Science, 11*(4), https://doi.org/10.1111/1467-8721.00182

Sharan, N. N., & Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon, 60*(8). https://doi.org/10.1016/j.heliyon.2020.e04572

Shoss, M. K. (2017). Job insecurity: An integrative review and agenda for future research. *Journal of Management, 43*(6), 1911-1939. https://doi.org/10.1177/0149206317691574

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422-445. https://doi.org/10.1037/1082-989X.7.4.422

Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal, 42*(9), 1600-1631. https://doi.org/10.1002/smj.3322

Vance, R. J., & Colella, A. (1990). Effects of two types of feedback on goal acceptance and personal goals. *Journal of Applied Psychology, 75*(1), 68-76. https://doi.org/10.1037/0021-9010.75.1.68

Yogeeswaran, K., & Dasgupta, N. (2014). The devil is in the details: Abstract versus concrete construals of multiculturalism differentially impact intergroup relations. *Journal of Personality and Social Psychology, 106*(5), 772-789 https://doi.org/10.1037/a0035830

Zhang, B., & Dafoe, A. (2019). Artificial intelligence: American attitudes and trends. *Available at SSRN 3312874.* https://doi.org/10.2139/ssrn.3312874

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies, 100,* 48-54. https://doi.org/10.1016/j.ijhcs.2016.12.008

투고일자 : 2024-08-20
수정일자 : 2024-11-12
게재일자 : 2024-12-17

한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2025. Vol. 38, No. 1, 1-26
https://doi.org/10.24230/kjiop.v38i1.1-26

Feedback Acceptance of AI-based Negative Feedback: The Mediating Effects of Threat and Perceived Credibility

Se Hee Cheon Seong Been In Ji Won Kang Jiyoung Park

Department of Psychology, Duksung Women's University

Artificial intelligence (AI) technology is rapidly advancing, leading to an increase in receiving AI-based feedback in the workplace. In this study, we explored how individuals respond to negative feedback from AI on creative tasks. Drawing upon feedback acceptance theory, we hypothesized that individuals would be less accepting AI-based negative feedback compared to feedback from humans. We expected that this effect would be mediated by higher levels of identity and realistic threats, and lower perceived source credibility, in the AI condition compared to the human condition. Furthermore, we hypothesized that the internal locus of control moderate the effects of these two threats on feedback acceptance. In an experimental study using Korean adults, we found that AI-based negative feedback increased both realistic and identity threats compared to human feedback. Contrary to our expectations, there were no significant differences in feedback acceptance or source credibility between the AI and human feedback conditions. While identity threat mediated the effect of AI-based feedback on lower feedback acceptance, realistic threat led to increased acceptance of AI feedback. Additionally, the internal locus of control moderated the relationship between identity threat and feedback acceptance. Based on these results, we discussed implications, limitations, and recommendations for future research.

Key words : Artificial intelligence, feedback acceptance, source credibility, identity threat, realistic threat, internal locus of control