



ISSN: 2288-7709

JEMM website: <https://accesson.kr/jemm>doi: <http://dx.doi.org/10.20482/jemm.2025.13.5.21>

Validation of Army Excellent Talent Characteristics Model using Virtual Data*

Cheol PARK¹

Received: September 10, 2025. Revised: September 25, 2025. Accepted: October 30, 2025.

Abstract

Purpose: This study aims to validate the potential of generative AI-based virtual data as a methodological alternative for testing talent models in military organizations, where data collection is restricted. **Methodology:** Using the Army Excellent Talent Characteristics Model as a framework, virtual data were generated through Conditional Tabular GAN (CTGAN) and evaluated using confirmatory factor analysis, independent sample t-tests, and multi-group structural equation modeling. **Results:** The first virtual dataset failed to meet convergent validity, but the regenerated dataset achieved acceptable construct reliability and average variance extracted values. The analysis showed that virtual data could replicate the overall measurement structure of actual data, yet differences emerged in the strength of structural paths. Specifically, leadership and communication satisfaction had stronger effects on trust in the actual dataset, while creativity was significant only in the virtual dataset. Trust influenced organizational citizenship behavior in both groups, but the explanatory power was higher in the real dataset. **Conclusion:** These findings suggest that virtual data can serve as a useful tool for validating latent structures, pre-testing survey designs, and enhancing reproducibility. However, its role should be considered complementary rather than a full substitute for real-world data when examining causal relationships in military human resource research.

Keywords : AI(Artificial Intelligence), Virtual Data, Excellent Talent Characteristics, Army, ChatGPT.

JEL Classification Code : C15, C52, J24, M12.

1. Introduction

오늘날 국방 분야는 빠르게 변화하는 안보 환경과 인구구조의 변화, 군 병력 자원의 질적 다양성 확대 요구 속에서 지속 가능한 인재 선발과 역량 기반 관리체계를 요구받고 있다. 특히, 육군은 장교 및 부사관 등 간부 중심의 정예화를 통해 전투력을 확보하고 있으며, 이에 따라 우수인재의 특성과 이들이 조직성과에 미치는 영향을 파악하는 것은 인적자원 전략의 핵심 과제 중 하나로 부상하고 있다.

그러나 기존의 실증연구들은 실제 수집된 데이터를 기반으

로 한 시점 단면자료(cross-sectional data)를 사용하거나 제한된 시계열 데이터를 분석하는 데 그치는 경우가 대부분이었다. 이로 인해 다음과 같은 세 가지 실무적 한계가 발생하고 있다. 첫째, 표본 수집에는 상당한 비용과 시간이 소요되며, 특히 군 조직처럼 폐쇄적이고 보안이 중시되는 환경에서는 설문조사 접근 자체가 어렵고 응답률 확보에도 제약이 따른다. 둘째, 수집 가능한 표본의 수가 제한적이기 때문에 통계적 검정력(statistical power)을 확보하기 어렵고, 표본 특성이 편향되어 연구의 일반화 가능성을 저해할 수 있다. 셋째, 무엇보다 군 인사정보는 민감 정보로 분류되어 있어 개인정보 보호

* This Work was supported by Jangan University Research Grant in 2025.

1 First Author. Assistant Professor, Department of Military Major, Jangan University, South Korea. Email: dreamatjau@jangan.ac.kr

© Copyright: The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited

법, 군 보안 규정, 연구윤리 가이드라인 등에 따라 외부 학술 연구자에게 제공되기 어렵다. 이러한 제약으로 인해 기존 연구 결과에 대한 재현 가능성(reproducibility)이 낮으며 반복적 분석을 통한 모형의 신뢰성 검증이나 보완 연구 또한 매우 제한적인 상황이다. 이처럼 군 조직의 우수인재 특성에 대한 연구는 현장성과 타당성 측면에서 중요성이 높음에도 불구하고, 현실적인 데이터 수집 한계로 인해 축적되기 어려운 구조적 한계를 지닌다.

이러한 배경 속에서 최근 급속히 발전하고 있는 생성형 인공지능(Generative AI)은 위와 같은 현실적 제약을 극복할 수 있는 대안으로 부상하고 있다. 생성형 AI는 기계학습 기반의 모델로 주어진 실제 데이터를 학습한 뒤 그 분포와 구조를 모사하는 가상의 새로운 데이터를 생성하는 기술이다. 이 중에서도 CTGAN(Conditional Tabular GAN)과 TimeGAN (Time-series Generative Adversarial Network)은 비정형이 아닌 정형(tabular) 데이터를 생성하는 데 최적화된 알고리즘으로 의료 임상데이터, 교육 평가데이터, 공공정책 데이터를 포함한 다양한 분야에서 실제 데이터와 통계적으로 유사한 특성을 유지하면서도 개인정보 유출 위험이 낮은 합성 데이터를 생성하는 데 사용되고 있다(Drechsler & Haensch, 2024). 예를 들어 CTGAN은 범주형 변수와 연속형 변수가 혼재된 데이터에서 각 변수 간 상호작용을 보존하며 샘플을 생성할 수 있다. 실제로 병원 환자 데이터를 바탕으로 만든 가상 환자군(synthetic cohort)이 진단 모형이나 예후 예측에서 실제 환자군과 거의 유사한 분석 결과를 보여준 바 있다(Kokosi & Harron, 2022). 이러한 기술은 개인정보 보호가 필수적인 영역에서 기존의 연구 방법론을 보완하고 보다 자유롭게 분석을 시도할 수 있는 기반을 제공해준다.

본 연구의 목적은 Bang(2012)이 연구에서 실제로 사용한 육군 우수인재 설문지를 기반으로 CTGAN을 활용해 가상 데이터를 생성하고, 이후 자기보고식 설문지의 한계를 반영하거나 성별, 계급, 연령 등 인구통계학적 변수를 1~2회 조정하여 모형의 통계적 유효성에 미치는 영향을 비교 분석함으로써 가상데이터의 활용 가능성을 검증하는 것이다. 이를 위해 먼저, 원자료를 기준으로 변수를 정의하고, 비교에 적합하게 분석 구조를 재구성하였다. 생성형 AI 모델은 OpenAI의 ChatGPT를 이용하여 가상데이터를 생성하였으며, 생성된 데이터와 실제 데이터를 통합한 후 확인적 요인분석을 통해 생성된 변수들에 대해 가상집단과 실제집단에 대한 독립표본 t-검정과 연구모형에 대한 회귀계수 비교 및 다집단 구조방정식 경로분석을 실시하였다. 마지막으로, 가상데이터에 문제

가 확인되는 경우나 결과 간의 차이가 있을 경우 생성조건을 더해 가상데이터를 재생성하고 반복 분석을 통해 타당성을 확보하고자 하였다.

본 연구를 통해 분석 결과가 유사할 경우 생성형 AI가 군 인재모형 검증에 있어 대체 가능한 도구로 활용될 수 있는지를 실증적으로 확인할 수 있을 것으로 기대하였다. 또한, 군 인적자원 분야에서 실증연구의 확장성과 재현 가능성을 확보할 수 있는 새로운 방법론적 대안을 제시하며, 향후 군 조직뿐만 아니라 공공정책, 조직행동, 마케팅 분야 등 다양한 사회과학 분야에서의 생성형 AI 기반 가상데이터 활용 가능성을 확인하는 계기가 될 것으로 기대하였다.

2. Literature Review

2.1. Characteristics of the Compared Paper

군 조직은 국가 안보라는 특수한 목적을 수행하는 집단으로 일관된 명령체계와 집단적 행동이 요구되는 계층적 구조를 갖고 있다. 이러한 특수한 환경 속에서 군 조직의 성과는 곧 전투력으로 직결되며, 이에 따라 구성원의 역량이 조직 전체의 효율성과 효과성에 미치는 영향이 매우 크다. 특히 장교 및 부사관과 같은 간부급 인력은 단순한 직무 수행자 이상의 전략적 의사결정자이자 부하들의 동기를 유발하는 리더로서 기능하므로, 이들의 '우수성'을 어떻게 정의하고 측정할 것인가는 군 조직의 인적자원관리에 있어 핵심적 문제로 제기되어 왔다.

Bang(2012)은 이러한 문제의식을 가지고 육군 조직 내 병사 및 부사관, 장교 등 간부의 우수인재 특성을 실증적으로 도출하고 이를 구조화된 인과모형으로 제시하였다. 구체적으로 그는 선행연구 고찰과 더불어 현직 군 간부 및 병사들을 대상으로 한 설문조사를 통해 군 조직에 적합한 우수인재의 특성요인을 도출하고 이를 통해 인재특성을 선행변수로, 신뢰를 매개변수로, 공정성을 조절변수로 구성하고 이들 요인이 조직시민행동에 미치는 영향과 관련된 구조화된 연구모형을 설계한 후 분석하였다. 또한, 그는 장교, 부사관, 병사라는 집단 간 인식 차이도 분석 대상으로 삼아, 각 인재 요인에 대한 집단별 중요도 인식 차이를 상관분석과 분산분석(ANOVA)을 통해 분석하였다. 그 결과, 각 계층 간 우수인재 요인에 대한 평가가 통계적으로 유의미하게 다르며, 이는 동일한 조직 내에서도 직책이나 역할에 따라 인재상에 대한 인식이 상이함을 실증하였다. 이 연구는 군 조직의 특수성을 고려한 실증적 인재모형을 제시하였다는 점에서 학문적·실무

적 기여가 크며, 이후 유사 조직(예: 경찰, 소방, 공공기관 등)에 대한 인재모형 연구에도 이론적 토대를 제공할 수 있다.

본 연구에서는 Bang(2012)의 연구모형을 기반으로 하되 구조방정식 다중집단분석을 통한 비교를 위해 조절효과와 관련된 공정성 변수를 제거하고 생성형 인공지능 기반 가상대이터를 활용하여 해당 인재모형의 통계적 유효성을 검증하고자 하였다.

2.2. Conceptual Framework of Excellent Talent Characteristics Factors

Bang(2012)이 제시한 군 조직 우수인재의 특성은 다음 일곱 가지 구성요인으로 정의되며, 각각은 이론적 선행연구와 군 조직 내 실증조사 결과를 통해 추출된 것이다. 각 요인의 개념은 다음과 같다.

Table 1: Characteristics of Excellent Talent

Characteristics	Definition	Reference
Leadership	From the perspective of organizational behavior, it refers to the process of fostering voluntary commitment and organizational innovation by articulating a vision, and the capability to exert influence through interactions with subordinates	Bass(1990), Yukl(1998), Jago(1982), Katz & Kahn (1978)
Trust	Trust refers to a belief that emerges from high-quality exchange relationships among members, facilitating interaction and cooperation. In military organizations, trust in superiors serves as a particularly critical factor	Deluga(1994), Cook & Wall (1980)
Communication Satisfaction	Communication satisfaction refers to members' level of satisfaction with communication-related aspects within the organization, including information sharing, feedback, and interaction. It is a multidimensional construct encompassing emotional responses, relational understanding, and the effectiveness of	Downs & Hazen (1977), Hecht(1978)

	communication	
Intrinsic Motivation	Motivation refers to an intrinsic interest in the activity itself and an internal energy that drives human behavior, affecting both the direction and persistence of actions	Deci & Ryan (2013), Woolfolk (1995), Berelson & Steiner (1964)
Creativity	Creativity refers to a cognitive and behavioral pattern of recognizing and solving problems in novel ways, leading to the creation of new outcomes. It represents both a process of generating innovation and an expression of self-actualization, shaped through the dynamic interaction between the individual and the environment	Bruner(1962); Rogers(1961); Torrance(1977); Taylor(1988)
Ethical Awareness	Ethical awareness refers to the disposition to act in accordance with moral principles such as honesty, fairness, and responsibility. It affects individuals' judgments and behaviors in situations of organizational conflict and is closely associated with trust and leadership	Schlenker & Forsyth (1977)
Equity	Equity refers to the belief that resources, opportunities, evaluations, and rewards are allocated fairly, affecting the formation of organizational trust and employee commitment	Greenberg(1990); Leventhal(1980)

2.3. Overview of Generative AI and Virtual Data Technologies

2.3.1. The Concept of Generative Artificial Intelligence

생성형 AI(Generative Artificial Intelligence)는 기존 데이터 또는 설계 구조를 학습한 후 해당 분포와 논리를 모사하여 새로운 데이터를 생성하는 인공지능 기술이다. 전통적인 분석은 관측된 데이터를 해석하거나 분류하는 데 중점을 두는 반면 생성형 AI는 데이터 자체를 만들어내는 데 초점을 둔다는 점에서 본질적으로 다르다. 이는 최근 자연어처리(NLP), 이미지 생성, 음성합성 등을 넘어, 구조화된 데이터나 변수 설계

에 기반한 '표 형식의 가상데이터' 생성으로 응용 범위를 넓히고 있다.

2.3.2. Concept and Construction Principles of Virtual Data

본 연구에서 사용하는 '가상데이터(virtual data)'란 실존하는 관측 데이터를 기반으로 일부 확장하거나 변형한 '합성데이터(synthetic data)'와는 달리 실제 데이터를 전혀 활용하지 않고도 이론적으로 정의된 모형 설계 기반(logic-based) 아래 생성된 데이터를 뜻한다. 이는 군 조직 인재모형 연구와 같이 데이터 보안성이 높은 영역에서 연구 반복성과 확장성을 확보하기 위한 새로운 접근 방식이다. 가상데이터는 변수 구성 및 문항 수, 리커트 5점 척도 등 측정 척도, 인과모형, 상관 구조 등 변수 간 관계 구조, 성별 비율, 직책별 비율 등 인구 통계적 설정값 등의 요소들을 기반으로 생성된다.

이러한 구조화된 설계를 입력값으로 제공하면, 생성형 AI 모델은 해당 구조를 만족하는 가상의 사례들을 생성하며, 이는 일종의 통계적 시뮬레이션(statistical simulation)의 범주에 속하면서도, 기계학습의 조건화된 생성을 활용해 데이터 현실감을 높이는 방향으로 진화하고 있다.

본 연구에서는 '가상데이터(virtual data)'를 이론적 모형 구조와 설정값을 기반으로 생성형 AI가 생성한 데이터로, 실제 데이터를 직접 학습하지 않는 것으로 정의하였으며, 기존의 실제 데이터를 학습한 후 그 통계적 분포를 바탕으로 생성한 데이터(GAN, VAE 등 활용)인 합성데이터(synthetic data)'와 이름을 명확히 구분하였다.

2.3.3. Generative AI Technologies for Virtual Data Generation

기존의 구조화된 데이터 생성에는 주로 GAN 기반 기술이 활용되며, 대표적으로 CTGAN(Conditional Tabular GAN), TableGAN, TVAE(Tabular Variational Autoencoder) 등이 있다. 그러나 이들은 기본적으로 실제 데이터를 기반으로 그 분포를 학습한 후 새로운 데이터를 생성하는 합성데이터 방식이기 때문에 완전한 가상데이터 생성에는 한계가 있다.

이에 따라 본 연구에서는 기존 문항 구조와 변수 설정을 Bang(2012)의 실제 데이터와 동일한 기준으로 설문지를 제공한 후 생성형 AI가 먼저 탐색하여 도출한 대상 집단의 특성에 기반을 둔 응답값을 정규분포, 다항분포, 계층 구조 등 설정된 규칙을 통해 자동 생성하는 방향으로 진행된다. Wu et al. (2022)은 화학반응 예측에서 반응식 규칙 기반의 가상데이터를 생성하고 이를 훈련에 활용해 기존 데이터 대비 예측력을 20% 이상 향상시킨 바 있으며, 이는 단순한 노이즈 부가가

아니라 설계 기반 데이터 생성이 유효함을 시사한다. 한편, AI 기반의 시뮬레이션 프레임워크인 VAMBN(VAE Modular Bayesian Network) 또는 TimeGAN과 같이 시계열 구조까지 반영할 수 있는 생성형 모델은 응답의 일관성과 분산을 조절하는 데 유용하며 변수 간 인과성까지 반영할 수 있는 잠재 가능성이 있다.

2.3.4. The Value and Empirical Need for Utilizing Virtual Data

가상데이터는 특히 다음과 같은 점에서 연구 도구로서 높은 잠재 가치를 지닌다. 첫째, 실제 데이터 없이도 특정 이론 모형의 분석 가능성과 통계적 적합성을 사전에 검증할 수 있다. 둘째, 조직 내 보안 문제로 인해 접근이 어려운 인적자원 데이터를 재현성 있게 대체할 수 있다. 셋째, 이론검증뿐 아니라 교육 및 분석 도구 개발에도 반복적으로 활용이 가능하다.

Kokosi & Harron(2022)은 완전한 가상데이터도 유의미한 분석 도구로 활용될 수 있으며 특히 인구집단 내 분포 구조나 변수 간 관계가 잘 정의되어 있는 경우 합성 데이터보다 오히려 더 높은 설명력을 보일 수 있다고 주장하였다. 다만, 이러한 가상데이터의 효용을 객관적으로 평가하기 위해서는 실제 데이터와 동일한 분석 모형을 적용한 후 결과의 통계적 차이가 유의미하지 않음을 입증할 필요가 있다. 따라서 본 연구는 이러한 실증적 접근을 통해 가상데이터의 연구 타당성을 평가하고자 하는 것이다.

3. Research Methods and Materials

3.1. Overview of the Study

본 연구는 연구자의 실제데이터 사용을 허락받아 군 조직에 적합한 우수인재의 특성요인을 도출한 Bang(2012)의 연구모형을 기반으로 하여 생성형 인공지능을 활용한 가상데이터를 통해 해당 모형의 통계적 구조가 재현 가능한지를 검증하고자 하였다. 이를 위해 기존의 응답 데이터를 직접 투입하지 않고 Bang(2012)의 설문 문항 구성 및 변수 간 인과구조를 설계 기반으로 모델링한 후 생성형 AI 기법을 활용해 동일한 구조를 갖는 가상데이터를 생성한다. 이후 원래의 분석 방식과 동일한 인과관계모형의 경로계수와 구조방정식 다중집단분석을 적용하여 분석한 결과와 Bang(2012)의 실증 분석 결과 간 유의미한 차이가 없을 경우 가상데이터를 활용한 군 인재모형 검증의 가능성과 유효성을 실증적으로 확인할 수 있다.

3.2. Research Design and Data Generation

본 연구는 다음과 같은 절차에 따라 설계되었다. 첫째, 비교 논문에서 제시한 설문지 문항의 구조, 문항 수, 측정 척도를 동일하게 적용하여 데이터 생성을 위한 기본 틀을 설계하였다. 이때, 생성형 AI 모델로는 조건부 생성적 적대 신경망(Conditional GAN) 계열 중 구조화된 표형 데이터에 특화된 CTGAN(Conditional Tabular GAN)을 선택하였다. 이 모델은 변수 간 상호작용과 분포 특성을 효과적으로 학습하며, 범주형 변수와 연속형 변수가 혼합된 데이터에서도 우수한 성능을 보인다는 점에서 본 연구 목적에 부합한다. 둘째, 실제 응답 데이터를 활용하지 않고 기존에 사용한 설문지에 제시된 성별, 계급, 연령 등 인구통계 특성을 기반으로 CTGAN에 입력할 기초 학습데이터 틀(seed structure)을 구성하였다. 이후 학습된 생성모델을 활용하여 비교 데이터와 동일한 응답자 수준의 가상 응답 데이터를 생성하였다. 이는 비교 논문에서 제시된 실제 분석 표본 수와 유사한 수준으로 설정하여 분석의 비교 가능성을 높이기 위함이다. 셋째, 생성된 가상데이터는 확인적 요인분석(CFA)을 통해 수렴타당도가 확보되면 가상데이터와 실제데이터를 집단으로 구분한 통합된 데이터로 구성하고, 다시 확인적 요인분석을 실시하여 변수를 생성한 후 독립표본 t-검정을 실시하였으며 구조방정식 모형 분석(SEM) 결과 경로계수의 방향성과 통계적 유의성 등을 비교 논문과 비교하였다. 이를 통해 가상데이터가 실제 측정 데이터를 얼마나 잘 모사(replicate)하는지를 실증적으로 검토하고자 하였다.

생성형 AI를 통한 설문 응답 데이터 생성을 위해 OpenAI의 GPT-4-turbo 모델을 활용하였으며, 문항별 응답 생성을 위한 temperature는 0.7로 설정하였다. 프롬프트는 "당신은 20대 남성 장교입니다. 다음 문항에 대해 5점 척도로 응답하십시오..."와 같은 방식으로 구성되었으며, 각 문항당 3개의 응답을 생성하여 평균값을 사용하였다. CTGAN은 에폭수 300, 배치 크기 64, 로그 변환된 입력 구조를 기반으로 학습되었고, 생성 시 성별/계급/연령 변수는 원자료의 분포 비율을 유지하도록 제약 조건이 부여되었다.

3.3. Structure of the Questionnaire

설문지는 Bang(2012)의 연구에서 사용된 인재 특성요인 모형을 바탕으로 구성되었으며, 총 7개의 하위요인으로 이루어져 있다. 본 연구에서는 해당 논문의 문항 수와 구성체계를 그대로 반영하였으며 조직시민행동(7점 척도)를 제외한 각

문항은 5점 리커트 척도(1 = 전혀 그렇지 않다 ~ 5 = 매우 그렇다)로 응답하도록 설계되었다. 각 하위요인의 문항 구성은 다음과 같다.

먼저 리더십은 Graen & Uhl-Bien(1995)의 연구를 근거로 "귀하의 상관은 귀하가 하는 일에 대해 어느 정도 만족하고 있다고 생각합니까?"와 같은 7개 문항을 측정하였다. 신뢰는 Lewicki & Bunker(1996)의 척도를 바탕으로 구성되었으며, "나는 나의 상관을 충분히 신뢰하고 있기 때문에 그(그녀)의 결정을 옹호하고 정당하다고 인정할 것이다."와 같은 문항을 포함하여 총 15문항으로 구성되었다. 커뮤니케이션 만족은 Downs & Hazen(1977)의 척도에 기반하여 "우리 조직은 나의 생각이나 느낌들을 솔직하게 표현할 수 있도록 해준다." 등의 총 24문항으로 측정되었으며, 동기는 Guay et al.(2000)의 Situational Motivation Scale(SIMS)을 참고하여 "나는 흥미 있는 일은 꼭 해내고야 만다." 등 총 6문항으로 구성되었다. 창의성은 Jung(2002)의 연구에서 개발된 척도를 활용하여 "나는 새로운 기능의 제품을 보면 사고 싶어 한다."와 같은 총 36개 문항으로 측정하였다. 윤리의식은 Schlenker & Forsyth(1977)의 연구를 기반으로 "나는 윤리적인 판단 기준을 갖고 행동하려고 노력한다." 등 총 10문항으로 측정하였다. 마지막으로 조직시민행동은 Organ(1988)의 OCB 척도에서 문항을 선별하여 "나는 조직의 발전을 위해 자발적으로 행동하는 편이다." 등의 총 12문항으로 측정하였다.

3.4. Procedures for Virtual Data Generation and Adjustment

본 연구에서 가상데이터를 활용하고자 하는 중요한 이유 중 하나가 군 조직과 같이 민감한 인적자원 데이터에 대한 접근 제약을 극복하면서도 실측 데이터를 대체할 수 있는 타당한 분석 자료를 확보하기 위함임은 앞에서 밝힌 바와 같다. 따라서 가상데이터를 생성한 후 1차 분석 결과를 검토하고 조정하는 절차는 <그림 1>과 같다.

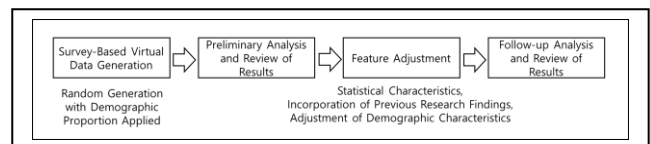


Figure 1: Virtual Data Generation and Adjustment Procedure

<그림 1>에 제시한 것처럼 먼저, 비교 논문에서 제시된

설문지를 기반으로 문항 수, 척도, 변수 구조 등을 동일하게 설정한 후 무작위(random) 방식으로 가상데이터를 생성한다. 이렇게 생성된 초기 데이터는 1차 분석에 투입되며 분석 결과는 실제 연구에서 보고된 결과와 비교하여 기본적인 적합성 여부를 검토한다. 이후 선행연구에서 확인된 신분, 성별, 연령, 학력 등 인구통계적 특성을 반영할 수 있도록 데이터를 조정한다. 이러한 조정 과정을 통해 가상데이터는 단순히 임의로 생성된 수준을 넘어 실제 표본의 인구통계학적 구조와 유사한 특성을 갖도록 설계된다. 마지막으로 조정된 가상데이터를 활용하여 2차 분석을 실시하고 그 결과를 검토함으로써 가상데이터가 실제 연구 맥락에서 타당하게 활용될 수 있는지를 단계적으로 확인한다.

3.5. Research Model

본 연구에서는 Bang(2012)이 제안한 인재 특성모형을 기초로 공정성의 조절효과를 제외한 수정된 연구모형을 비교 분석에 활용하였다. 전체 모형은 아래 <그림 2>와 같다.

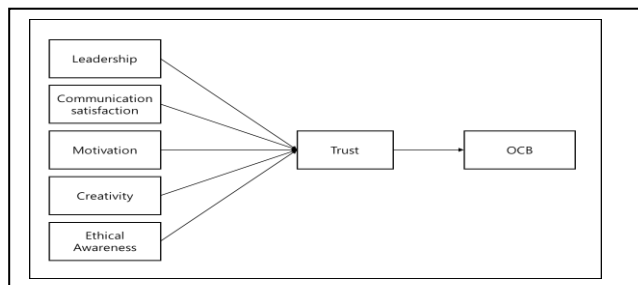


Figure 2: Research Model

3.6. Hypothesis Development

본 연구가 비교 논문에서 실측 데이터를 기반으로 분석된 군 조직의 우수인재 특성모형을 적용하여 생성형 AI 기반 가상데이터를 통해 통계적으로 재현할 수 있는지를 검증하는데 목적이 있음은 앞에서 밝힌 바와 같다. 특히 생성된 가상데이터의 인과관계 분석 가능성을 확인하며, 각 변수 간 경로 구조뿐 아니라 변수 수준(평균값)에서도 실측 데이터와 가상데이터 간 유사성을 비교 분석하고자 한다. 이를 위해 다음과 같은 가설을 설정하였다.

가설 1 (H1): 생성형 AI 기반 가상데이터의 측정모형은 수렴 타당도와 내적일관성을 충족할 것이다.

가설 2 (H2): 가상데이터로 생성한 리더십, 윤리의식 등 각 변수의 평균값은 비교 논문의 실측 기반 분석 결과와 통계적으로 유의미한 차이를 보이지 않을 것이다.

가설 3 (H3): 가상데이터와 실제데이터 집단으로 구분한 구조방정식 다중집단분석 결과 두 집단은 통계적으로 유의미한 차이를 보이지 않을 것이다

4. Empirical Analysis

4.1. Virtual Data Generation (Phase 1)

확인적 요인분석(CFA)을 통해 측정모형의 적합성을 검증하고자 하였다. 그러나 랜덤 생성된 1차 가상데이터는 표준적 재치가 0.5 미만인 문항이 다수 발견되어 수렴타당도를 확보할 수 없었다. 확인적 요인분석을 통해 수렴타당도가 확보되지 못하는 경우에는 문항 제거를 통한 변수화가 불가능하다. 따라서 1차 가상데이터는 인과관계 분석에 적합한 데이터가 아니라는 결론을 내렸다. 최초 ChatGPT에 설문지를 업로드한 후 가상 데이터 생성을 위한 프롬프트는 아래와 같다.

첨부한 파일은 가상데이터 생성을 위한 설문지이며, 첨부한 이미지는 실제 분석된 인구통계적 특성에 대한 결과 값입니다. 데이터는 사병, 부사관, 위관장교, 영관장교로 구분되어야 하며 이들의 특성은 검색을 통해 도출한 후 반영해야 합니다. 인구통계적 특성 비율에 적합하게 첨부파일의 각 문항에 대해 가상 데이터를 생성해주기 바랍니다.

4.2. Virtual Data Generation (Phase 2)

따라서 <그림 1>의 절차와 같이 2차 데이터는 통계적 특성, 자기보고식 설문지의 한계로 제시되고 있는 사회적 바람직성 등을 고려해 재생성하였다. 재생성을 위한 프롬프트는 다음과 같이 구성하였다. ChatGPT 프롬프트에는 아래와 같이 자기보고식 설문 문항의 분포 특성(평균 중심, 표준편차 1 이내)을 유도하는 조건과, "사회적 바람직성"을 고려한 응답 분포를 요구하는 문장을 삽입하였다.

첨부한 파일은 가상데이터 생성을 위한 설문지이며, 첨부한 이미지는 실제 분석된 인구통계적 특성에 대한 결과 값입니다. 데이터는 사병, 부사관, 위관장교, 영관장교로 구분되어야

하며 이들의 특성은 검색을 통해 도출한 후 반영해야 합니다. 자기보고식 설문지의 경우 중앙값을 기준으로 표준오차가 1 미만으로 응답되는 것이 일반적이며, 사회적 바람직성을 고려하여 응답되어야 합니다. 선행변수들과 신뢰, 신뢰와 조직시민 행동 간의 관계, 인과관계 결과에 대한 선행연구들을 검색하여 이 내용을 반영하고, 인구통계적 특성 비율에 적합하게 첨부파일의 각 문항에 대해 가상 데이터를 생성해주기 바랍니다.

4.2.1. Validation of the Measurement Model

먼저 <가설 1>의 검증을 위해 재생성된 데이터를 활용하여 확인적 요인분석을 실시한 결과 개념신뢰도는 0.903~0.981, 분산추출지수는 0.571~0.622로 나타났으며 내적일관성을 검증한 Cronbach's α 계수도 0.838~0.965로 확인되었다. 이를 통해 본 연구에서 사용된 측정모형은 수렴타당도와 신뢰도가 확보되었다고 결론내렸다. 따라서 <가설 1>은 채택되었다.

Table 2: Confirmatory Factor Analysis Results for the Virtual Data (n=697)

Variable	Initial Item	Final Item	C.R	AVE	Cronbach's α
Leadership	7	7	0.903	0.572	0.838
Communication Satisfaction	24	24	0.972	0.595	0.951
Motivation	6	6	0.908	0.622	0.840
Creativity	34	34	0.981	0.584	0.965
Ethical Awareness	10	10	0.930	0.571	0.882
Trust	15	15	0.956	0.592	0.923
OCB	12	12	0.943	0.580	0.905

<표 3>에 제시한 것처럼 통합데이터에 대한 최초 확인적 요인분석 결과 변수별로 각 문항의 요인적재치가 기준치(0.5)를 충족하지 못한 커뮤니케이션 만족의 1개 문항, 창의성의 9개 문항, 윤리의식의 2개 문항, 조직시민행동의 1개 문항을 통제 후 확인적 요인분석을 실시한 결과 개념신뢰도는 0.874~0.970, 분산추출지수는 0.506~0.656으로 나타났으며 내적일관성을 검증한 Cronbach's α 계수도 0.838~0.924로 확인되었다. 이를 통해 본 연구에서 사용된 측정모형은 수렴타당도와 신뢰도가 확보되었다고 결론내렸다.

Table 3: Confirmatory Factor Analysis Results for the Combined Data (n=1,394)

Variable	Initial Item	Final Item	C.R	AVE	Cronbach's α
Leadership	7	7	0.896	0.551	0.856
Communication	24	23	0.970	0.583	0.924

Satisfaction					
Motivation	6	6	0.874	0.537	0.914
Creativity	34	23	0.959	0.508	0.794
Ethical Awareness	10	8	0.916	0.579	0.868
Trust	15	15	0.966	0.656	0.936
OCB	12	11	0.918	0.506	0.838

4.2.2. Correlation Analysis Results

가상데이터의 상관계수는 0.281~0.489($P < .01$)로 나타났으며, 실제데이터의 상관계수는 0.327~0.774($P < .01$)로 나타나 가상데이터와 실제데이터의 변수 간의 관계와 방향성은 동일한 것으로 확인되었다(Appendix 1, 2 참고). 그러나 가상데이터에 비해 실제데이터의 상관계수가 전반적으로 높은 것으로 나타났으며 표준편차 또한 가상데이터(0.17~0.24)에 비해 실제데이터(0.48~1.24)의 폭이 큰 것으로 확인되었다. 인과관계분석이 가능한지에 대한 확인과 판별타당도를 확인하기 위해 통합 데이터에 대한 상관관계분석 결과 각 변수는 양(+)의 관계가 있는 것으로 확인되었으며, 분산추출지수(AVE)와 상관계수의 제곱 값인 결정계수(r^2)를 비교한 결과 분산추출지수 값이 높아 판별타당도를 확보한 것으로 결론내릴 수 있었다(Appendix 3 참고).

4.2.3. Independent Samples t-Test Results

<가설 2>의 검증을 위해 실제 분석 결과와 가상데이터의 2개 집단에 대한 독립표본 t-검정 결과 커뮤니케이션 만족만이 통계적으로 유의한 차이가 없었다. 그러나 리더십은 가상데이터가 실제데이터보다 통계적으로 유의하게 높은 것으로 나타났으며, 동기, 창의성, 신뢰, 윤리의식, 조직시민행동은 실제데이터의 평균이 통계적으로 유의한 것으로 나타났다. 구체적으로 리더십은 가상데이터에서 유의하게 높게 나타난 반면, 대부분의 변수는 실제데이터에서 평균값이 더 높았으며, 특히 동기 및 조직시민행동의 차이가 크게 나타났다. 따라서 <가설 2>는 기각되었다(Appendix 4 참고).

4.2.4. Multi-Group Analysis of Path Coefficients

<가설 3>을 검증하기 위해 가상데이터와 실제 데이터를 통합하고 집단변수(1: 가상, 2: 실제)를 기준으로 구조방정식 다중집단 분석을 실시하였다.

분석 결과, 리더십과 커뮤니케이션 만족은 실제데이터 집단에서 신뢰에 미치는 영향이 가상데이터 집단보다 유의하게 높았다($Z = -4.06, p < .001$; $Z = -3.45, p < .001$). 반대로 창의성은 가상 집단에서만 신뢰에 유의한 영향을 미쳤으며 두 집단 간 차이도 통계적으로 유의하였다($Z = 2.05, p < .05$). 반면, 동기와 윤리의식은 두 집단 간 차이가 나타나지

않았다. 설명력(R^2)을 살펴보면 신뢰에 미치는 영향은 가상데이터 집단이 0.351, 실제데이터 집단이 0.689로 나타나 실제데이터가 신뢰 변인을 더 잘 설명하는 것으로 확인되었다. 한편, 신뢰는 두 집단 모두에서 조직시민행동에 유의한 영향을 미쳤다. 그러나 실제데이터 집단의 경로계수($b=0.762$, $SE=0.041$)가 가상데이터 집단($b=0.485$, $SE=0.043$)보다 유의하게 높았으며($Z=-4.66$, $p<.001$). 설명력(R^2) 또한 실제데이터 집단(0.436)이 가상데이터 집단(0.142)보다 높았다. 구체적으로 리더십, 커뮤니케이션 만족, 신뢰는 실제집단에서 경로계수가 더 강하게 나타났으며, 창의성은 가상집단에서만 유의하였다. 이는 변수 간 관계 강도는 실제와 다를 수 있음을 시사한다. 집단별 경로계수에 대한 비교 결과는 <표 4>에 제시하였다.

Table 4: Comparison of Path Coefficients by Group

Direct Effect	Virtual Group b (SE)	Actual Group b (SE)	Z	p
Leadership → Trust	0.059 (0.030)	0.250 (0.037)	-4.06	<.001
Communication Satisfaction → Trust	0.290 (0.040)	0.533 (0.058)	-3.45	<.001
Motivation → Trust	0.113 (0.027)	0.124 (0.026)	-0.3	n.s.
Creativity → Trust	0.181 (0.045)	0.052 (0.044)	2.05	<.05
Ethical Awareness → Trust	0.173 (0.029)	0.105 (0.032)	1.57	n.s.
Trust → OCB	0.485 (0.043)	0.762 (0.041)	-4.66	<.001

Note) n.s. = not significant

집단 간 경로 동일성을 검증하기 위해 집단별로 경로를 자유 추정한 자유모형과 집단 간 동일성을 제약한 제약모형의 χ^2 차이 검정을 실시한 결과 리더십과 커뮤니케이션 만족, 창의성, 그리고 신뢰와 조직시민행동 간의 경로에서는 집단 간 차이가 유의하여 동일성 제약이 기각되었다. 반면, 동기와 윤리의식, 윤리의식과 신뢰는 동일성 제약이 기각되지 않아 두 집단 간 차이가 없는 것으로 확인되었다. ΔCFI 역시 동일한 패턴을 보였다. 따라서 <가설 3>은 기각되었다. 분석 결과는 <표 5>에 제시하였다.

Table 5: Equality Constraint Test by Path (χ^2 Difference, ΔCFI)

Path	$\Delta\chi^2(df=1)$	p	ΔCFI	Decision
Leadership → Trust	16.4	<.001	0.03	Invariance Rejected

Communication Satisfaction → Trust	11.9	<.001	0.02	Invariance Rejected
Motivation → Trust	0.1	n.s.	0.002	Invariance Accepted
Creativity → Trust	4.2	<.05	0.01	Invariance Rejected
Ethical Awareness → Trust	2.5	n.s.	0.005	Invariance Accepted
Trust → OCB	21.7	<.001	0.04	Invariance Rejected

이상의 결과는 생성형 AI 기반 가상데이터가 실제데이터의 구조적 관계를 상당 부분 재현할 수 있음을 보여준다. 그러나 리더십, 커뮤니케이션 만족, 신뢰와 조직시민행동 간의 경로에서는 실제데이터와 가상데이터 간 차이가 존재하였다. 특히 실제데이터에서는 리더십과 커뮤니케이션 만족이 신뢰에 미치는 영향이 더 강하게 나타났으며, 신뢰가 조직시민행동에 미치는 영향 역시 더 크게 나타났다. 반대로 창의성은 가상데이터에서만 신뢰에 유의하게 영향을 미치는 것으로 나타났다. 이러한 결과는 가상데이터가 전체적인 구조는 모사하였으나 변수별 영향력의 강도까지는 완전히 재현하지 못했음을 의미한다.

5. Results and Discussion

본 연구는 육군 우수인재 특성모형을 대상으로 생성형 AI 기반 가상데이터가 실제데이터와 어느 수준까지 동등성을 달성하는지 다층적으로 검증하였다. 1차 가상데이터는 수렴타당도를 충족하지 못해 제외되었고, 재생성한 2차 가상데이터는 확인적 요인분석에서 개념신뢰도 0.903~0.981, 분산추출지수 0.571~0.622로 측정 일관성을 확보하였다. 측정모형 수준의 유효성은 지지되었으나 독립표본 t-검정에서 일부 잠재변수와 지표의 평균차가 관찰되었고, 다중집단 구조방정식에서는 리더십→신뢰($Z=-4.06$), 의사소통 만족→신뢰($Z=-3.45$), 신뢰→조직시민행동($b=0.762$ vs. 0.485 , $Z=-4.66$)이 실제집단에서 더 강하게 나타났다. 창의성→신뢰는 가상집단에서만 유의하며 두 집단 간 차이도 유의하였다($Z=2.05$). 창의성 → 신뢰 경로가 가상데이터 집단에서만 유의하게 나타난 점은, LLM 기반 생성 방식이 문항 간 상관구조를 맥락 없이 기계적으로 연결할 가능성이 있음을 시사한다. 반면 실제데이터에서는 해당 관계가 조직 맥락 내 문화적 규범이나 역할 기대 등 맥락 변수에 의해 제약 받았을 가능성이 크다. 이는 생성형 AI 기반 가상데이터가 구조모형의 골격은 모사하더라도, 관계

강도나 맥락적 민감도는 충분히 반영하지 못할 수 있다는 점을 보여준다.

설명력 역시 실제집단이 신뢰 0.689 대 0.351, 조직시민행동 0.436 대 0.142로 높았다. 이런 결과를 볼 때 가상데이터는 잠재구조와 요인 패턴 등 모형의 큰 골격은 상당 부분 재현했지만 구조경로의 강도와 일부 평균·분산 특성은 충분히 보존하지 못한 것을 알 수 있다. 따라서 본 연구는 완벽한 동등성에는 이르지 못했음을 분명히 하면서도 생성형 AI 기반 가상데이터가 측정 수준에서의 구조 모사는 가능하다는 점과 구조적 관계의 정밀 보존이 후속 과제임을 함께 보여준다.

이러한 결과는 가상데이터 활용을 실제 자료의 전면적 대체로 이해하기보다는 설문 설계의 사전 점검, 모형 민감도 분석, 재현 가능성 강화를 위한 외부 공유 자료 등 보완적·점증적 용도부터 적용해야 한다는 점을 시사한다. 동시에 다수 가상데이터 세트를 결합 규칙으로 통합해 분석하는 절차, 공분산 구조를 학습하는 표형 특화 생성기의 활용, 작업지향 유틸리티 검증의 병행 등 방법론적 개선을 통해 구조경로 보존과 집단 간 동질성 확보 가능성을 점차 높일 필요가 있다 (El Emam et al., 2024; Xu et al., 2019; Kotelnikov et al., 2023; Yoon et al., 2019).

본 연구의 결과는 생성형 AI 기반 가상데이터가 측정 수준에서는 실제데이터의 구조를 상당 부분 재현한다는 가능성을 보여준다는 점에서 그 의의가 있다. 구체적으로 2차로 생성한 가상데이터에서 개념신뢰도와 분산추출지수가 충분히 확보되었다는 사실은 잠재구조와 요인패턴이 비교적 안정적으로 묘사된다는 점을 시사한다. 동시에 다중집단 구조방정식에서 경로계수와 설명력이 의미 있게 이탈했다는 사실은 가상데이터가 이론적 관계의 강도와 방향성까지 동일하게 보존하기에는 아직 제약이 크다는 점을 드러낸다. 이런 결과를 볼 때 학술 연구에서는 가상데이터를 이론검증의 최종 근거로 사용하기보다 설문지 구성 타당화와 모형의 식별성 점검, 측정오류 민감도 확인 등 전처리 단계의 과학적 도구로 위치시키는 편이 타당하다고 할 것이다. 특히 본 연구에서 신뢰의 결정요인에 대한 경로 강도가 실제데이터 집단에서 체계적으로 더 컸고 가상데이터 집단에서는 창의성과 신뢰 간의 관계에서만 유의하게 나타났다는 결과는 가상데이터가 맥락적 제약과 규범적 신호를 충분히 반영하지 못할 때 구조적 추론이 왜곡될 수 있음을 경고한다. 따라서 향후 유사 연구는 측정동등성 확보를 출발점으로 삼되, 경로계수와 설명력의 허용오차 범위를 사전에 정의해 가상데이터의 연구용 적

합성을 단계적으로 판단하는 절차가 요구된다.

실무 측면에서 본 연구의 결과는 가상데이터를 활용할 때 용도와 한계를 분명히 구획해야 함을 보여준다. 실제데이터 집단에서 리더십과 커뮤니케이션 만족이 신뢰에 미치는 영향이 더 강했고, 신뢰가 조직시민행동을 설명하는 힘도 현저히 컸다는 사실은 군 조직의 위계·규범·명령체계 같은 맥락 변수가 신뢰 형성과 역할 외 행동에 실질적으로 작동하고 있음을 시사한다. 가상데이터는 이러한 맥락적 압력과 비대칭적 정보구조를 충분히 내재화하지 못해 경로 강도를 과소추정하거나 대체 경로를 과대추정할 가능성이 있다. 따라서 교육훈련 설계나 선발·배치 정책과 같이 구조적 관계의 강도를 전제로 하는 의사결정에는 실제데이터 기반 추정치를 기준선으로 두고, 가상데이터는 항목 난이도 조정, 천정·바닥 효과 사전 점검, 희귀 표본 시나리오의 민감도 분석 등 보조적 용도로 제한하는 편이 안전하다. 특히 본 연구에서 가상데이터 집단에서만 유의했던 '창의성→신뢰'경로는, 창의성 고도화를 통해 신뢰를 높이려는 개입설계가 가상데이터만으로는 과대평가될 수 있음을 암시한다. 리더십 개발과 커뮤니케이션 개선이 신뢰와 조직시민행동에 미치는 실효성이 실제데이터에서 더 강하게 관찰된 만큼 실무 적용에서는 이 두 축을 우선순위로 두되 가상데이터는 프로그램 사전 모의와 문항 설계 검증에 한정하는 운영 원칙이 합리적이다.

정책 차원에서는 가상데이터의 사용 가능 범위와 금지 범위를 명확히 라벨링하는 거버넌스가 필요하다. 본 연구와 같은 결과가 축적될수록 가상데이터 공개 시 "측정 수준 검증·교육·외부협업" 과 같은 허용 용도와 "구조경로 추정·효과 크기 비교의 최종 근거" 와 같은 제한 용도를 함께 명시하는 것이 바람직하다. 외부 연구자와의 협업이나 감사 목적의 공유에서는 재현가능성 제고라는 장점이 크지만 정책 효과성 판단이나 인사정책의 개입 강도를 산정하는 단계에서는 실제데이터 기반 추정에 대한 교차검증을 의무화해야 한다. 또한 기관 내부에서는 가상데이터의 적합성 판단을 위한 체크리스트를 운영하여 평균·분산 정합성, 경로계수 근접성, 설명력 격차 등 핵심 지표가 사전 정의된 임계치를 충족할 때에만 제한적 확장을 허용하는 단계적 개방 전략이 요구된다.

본 연구가 다양한 관점의 시사점을 도출했지만 다음과 같은 한계점에 대해서는 향후 연구에서 보완되어야 할 것이다. 먼저, 본 연구의 경험은 가상데이터 생성 파이프라인의 정교화 방향을 구체적으로 제시해야 함을 지적하고 있다. 즉, 측정

수준의 동등성이 확보된 반면 구조수준의 이탈이 발생했다는 사실을 고려할 때 향후에는 조건부 생성과 규칙 기반 제약을 결합해 평균·분산·공분산의 위치·척도 정보를 사전적으로 부여하는 절차가 유효할 것이다. 또한 실제데이터 집단에서 관찰된 리더십·의사소통 만족의 강한 영향력을 기준으로 가상데이터 생성 단계에서 관련 지표 간 상관구조와 공동분산을 보존하려는 제약 조건을 도입하면 구조경로의 근접성을 높일 수 있을 것이다. 분석 단계에서도 단일 합성세트에 의존하기보다 다수의 재생성 세트를 활용해 추정치를 집계하고 경로계수와 설명력의 분산을 함께 보고하는 방식으로 불확실성을 투명하게 관리할 필요가 있다. 무엇보다 본 연구에서 드러난 창의성 경로의 과대·과소추정 가능성은 맥락 변수를 내재화하지 않은 가상데이터가 특정 관계를 체계적으로 왜곡할 수 있다는 점을 일깨운다. 따라서 후속 연구에서는 맥락 변수의 조작적 정의를 정교화하고, 현장 규범과 절차의 신호를 생성 과정에 내재화하는 방법을 탐색해야 한다. 특히 본 연구에서 가상데이터는 구조적 골격은 비교적 유사하게 재현하였으나, 경로 강도 및 설명력에서 차이를 보였다. 이는 정책 결정이나 프로그램 설계와 같이 구조 경로의 효과 크기가 중시되는 분야에서는 실제데이터 기반 추정을 기준선으로 설정해야 함을 시사한다. 가상데이터는 문항의 난이도, 분산, 설문 설계 적합성 등을 사전 점검하는 데 유용하므로, 실무에서는 보완적 도구로서 신중하게 활용되어야 할 것이다.

References

- Bang, W. P. (2012). The Effects of Characteristics of Excellent Manpower on the Organizational Performance in Army. (Doctorial dissertation, Anyang University). Retrieved June, 2012 from <https://www.riss.kr/link?id=T12934966>
- Bass, B. M. (1990). From transactional to transformational leadership: Learning to share the vision. *Organizational dynamics*, 18(3), 19-31.
- Berelson, B., & Steiner, G. A. (1964). Human behavior: An inventory of scientific findings. Harcourt, Brace & World.
- Bruner, J. S. (1962). The conditions of creativity. In Contemporary Approaches to Creative Thinking, 1958, University of Colorado, CO, US; This paper was presented at the aforementioned symposium.. Atherton Press.
- Cook, J., & Wall, T. (1980). New work attitude measures of trust, organizational commitment and personal need non-fulfilment. *Journal of occupational psychology*, 53(1), 39-52.
- Deci, E. L., & Ryan, R. M. (2013). Intrinsic motivation and self-determination in human behavior. Springer Science & Business Media.
- Deluga, R. J. (1994). Supervisor trust building, leader-member exchange and organizational citizenship behaviour. *Journal of occupational and Organizational Psychology*, 67(4), 315-326.
- Downs, C. W., & Hazen, M. D. (1977). A factor analytic study of communication satisfaction. *The Journal of Business Communication* (1973), 14(3), 63-73.
- Drechsler, J., & Haensch, A. C. (2024). 30 years of synthetic data. *Statistical Science*, 39(2), 221-242.
- El Emam, K., Mosquera, L., Fang, X., & El-Hussuna, A. (2024). An evaluation of the replicability of analyses using synthetic health data. *Scientific Reports*, 14(1), 6978.
- Graen, G. B., & Uhl-Bien, M. (1995). Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *The leadership quarterly*, 6(2), 219-247.
- Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of management*, 16(2), 399-432.
- Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and emotion*, 24(3), 175-213.
- Hecht, M. L. (1978). The conceptualization and measurement of interpersonal communication satisfaction. *Human Communication Research*, 4(3), 253-264.
- Jago, A. G. (1982). Leadership: Perspectives in theory and research. *Management science*, 28(3), 315-336.
- Jung, Y. E. (2002). The Relationship between Everyday Creativity, Individuation-Relatedness, and Psychological and Social Adjustment. (Doctorial dissertation, Korea University). Retrieved June, 2002 from <https://www.riss.kr/link?id=T8347015>
- Katz, D., & Kahn, R. L. (1978). The social psychology of organizations (Vol. 2, p. 528). New York: wiley.
- Kokosi, T., & Harron, K. (2022). Synthetic data in medical research. *BMJ Medicine*, 1(1), e000167. <https://doi.org/10.1136/bmjmed-2022-000167>
- Kotelnikov, A., Baranchuk, D., Rubachev, I., & Babenko, A. (2023, July). Tabddpm: Modelling tabular data with diffusion models. In International conference on machine learning (pp. 17564-17579). PMLR.
- Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. In Social exchange: Advances in theory and research (pp. 27-55). Boston, MA: Springer US.
- Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.), Trust in organizations: Frontiers of theory and research (pp. 114-139). Sage.
- Organ, D. W. (1988). Organizational citizenship behavior: The good soldier syndrome. Lexington Books.
- Rogers, C. R. (1995). On becoming a person: A therapist's view of psychotherapy. Houghton Mifflin Harcourt.
- Schlenker, B. R., & Forsyth, D. R. (1977). On the ethics of psychological research. *Journal of Experimental Social Psychology*, 13(4), 369-396.
- Taylor, I. A. (1988). Various approaches to and definitions of creativity. In R. J. Sternberg (Ed.), The nature of creativity: Contemporary psychological perspectives (pp. 99-121). Cambridge University Press.

- Torrance, E. P. (1977). Creativity in the classroom. Washington National Education Association.
- Woolfolk, A., & Margetts, K. (2012). Educational psychology Australian edition. Pearson Higher Education AU.
- Wu, X., Zhang, Y., Yu, J., Zhang, C., Qiao, H., Wu, Y., ... & Duan, H. (2022). Virtual data augmentation method for reaction prediction. *Scientific Reports*, 12, 17098. <https://doi.org/10.1038/s41598-022-21524-6>
- Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. arXiv:1907.00503.
- Yoon, J., Jarrett, D., & Van der Schaar, M. (2019). Time-series generative adversarial networks. *Advances in neural information processing systems*, 32.
- Yukl, G. A. (1998). Leadership in organizations (4th ed.). Prentice Hall.

Appendixes

Appendix 1: Correlation Analysis Results for the Virtual Data (n=697)

	1	2	3	4	5	6	7
1. Leadership	-						
2. Communication satisfaction	.399**	-					
3. Motivation	.304**	.363**	-				
4. Creativity	.320**	.429**	.267**	-			
5. Trust	.334**	.489**	.361**	.399**	-		
6. Ethical awareness	.319**	.393**	.284**	.379**	.427**	-	
7. OCB	.281**	.392**	.317**	.333**	.294**	.377**	-
Mean	3.77	3.66	3.85	3.41	3.66	3.76	5.17
Standard Deviation	0.23	0.17	0.24	0.15	0.19	0.22	0.24

Note: ** p<.01

Appendix 2: Correlation Analysis Results for the Real Data (n=697)

	1	2	3	4	5	6	7
1. Leadership	-						
2. Communication satisfaction	.684**	-					
3. Motivation	.512**	.622**	-				
4. Creativity	.327**	.354**	.424**	-			
5. Trust	.688**	.774**	.649**	.384**	-		
6. Ethical awareness	.388**	.470**	.505**	.394**	.505**	-	
7. OCB	.512**	.638**	.656**	.402**	.660**	.652**	-
Mean	3.40	3.46	4.91	3.64	3.63	3.92	4.93
Standard Deviation	0.80	0.64	1.24	0.48	0.80	0.75	0.92

Note: ** p<.01

Appendix 3: Correlation Analysis Results for the Combined Data (n=1,394)

	1	2	3	4	5	6	7
1. Leadership	(.551)						
2. Communication satisfaction	.682**	(.583)					
3. Motivation	.258**	.406**	(.537)				
4. Creativity	.205**	.271**	.493**	(.508)			
5. Trust	.642**	.746**	.534**	.358**	(.579)		
6. Ethical awareness	.318**	.419**	.492**	.414**	.491**	(.656)	
7. OCB	.519**	.636**	.449**	.316**	.638**	.584**	(.506)
Mean	3.58	3.56	4.38	3.52	3.65	3.84	5.05
Standard Deviation	0.61	0.48	1.04	0.37	0.58	0.56	0.69

Note: ** p<.01

Appendix 4: Results of Independent Samples t-Test Integrating the Second Virtual Dataset

Variables	Group	N	Mean	Standard Deviation	Mean Difference	t-value	p-value
Leadership	Virtual	697	3.49	0.53	0.089	2.459	0.014
	Actual	697	3.40	0.80			
Communication satisfaction	Virtual	697	3.49	0.52	0.028	0.910	0.363
	Actual	697	3.46	0.63			
Motivation	Virtual	697	3.51	0.54	-1.401	-27.264	0.000
	Actual	697	4.91	1.24			
Creativity	Virtual	697	3.51	0.51	-0.989	-3.706	0.000
	Actual	697	3.61	0.49			
Trust	Virtual	697	3.53	0.52	-0.099	-3.023	0.003
	Actual	697	3.63	0.80			
Ethical awareness	Virtual	697	3.47	0.53	-0.454	-12.976	0.000
	Actual	697	3.92	0.75			
OCB	Virtual	697	3.48	0.54	-1.448	-35.824	0.000
	Actual	697	4.93	0.92			