

AI 전환(AX) 진단을 위한 적응형 하네스 아키텍처: TIM 분류체계 기반 공통 코어와 플러그인 설계

최광훈*, 조상훈**

An Adaptive Harness Architecture for AI Transformation Diagnosis: A TIM-Taxonomy Core with Target and Stage-Specific Plug-Ins

Kwang hun Choi*, Sanghun Cho**

본 논문은 2026년도 한국과학기술정보연구원 주요사업(수요대응형지역R&D혁신지원체계구축, No. K26L4M2C3), 자체연구사업(정책지능 및 AX 인프라 구축 방안 연구, No. J26JR032-26)의 지원을 받아 수행한 결과입니다.

요 약

본 연구의 목적은 지역 AI 전환(AX) 정책이 요구하는 사업기획 평가와 기업 역량 진단을 동일 기준에서 처리하는 진단 아키텍처를 제시하는 것이다. 이를 위해 설계과학연구 방법론에 따라 적응형 하네스를 설계하였다. 하네스는 TIM 분류체계, 다중 AI 모델 합의 채점, 집계 규약으로 구성되는 고정 공통 코어와 그 위에 탑재되는 두 종류의 교체형 플러그인으로 이루어지며, 적응은 코어의 재설계가 아닌 플러그인 교체를 통해 수행된다. 제안 하네스를 기업 281개사와 지역 사업기획 40건에 적용한 결과, 단일 코어가 두 대상에서 유사한 상대 형태의 도메인 프로파일을 산출하면서도 각 대상의 고유 특성을 분별함을 확인하였다. 채점 코어의 평가자 간 신뢰도와 내용타당도가 이러한 결과를 뒷받침한다. 제안 구조는 코어 불변, 부품 적응, 공통 좌표의 세 설계 원칙으로 정리되며, 진단 산출은 의사결정을 지원하는 용도로 한정된다.

Abstract

This study presents an adaptive diagnostic architecture that processes the two diagnoses required by regional AI-transformation (AX) policy, project-plan evaluation and firm-capability diagnosis, on a single standard. Following the design science research methodology, we develop an adaptive harness. The harness consists of a fixed common core and two kinds of swappable plug-ins adaptation is achieved by swapping plug-ins rather than redesigning the core. Applied to 281 firms and 40 regional project plans, a single core produced similar relative domain profiles across the two targets while differentiating the distinct traits of each. The reliability and content validity of the scoring core support these results. The design is summarized in three principles, a fixed core, part-level adaptation, and shared coordinates, and the outputs are positioned as decision support.

Keywords

AI transformation, harness architecture, TIM taxonomy, adaptive diagnosis, maturity model, design science research

* 한국과학기술정보연구원 지역혁신연구센터장

- ORCID: <https://orcid.org/0000-0001-5791-1129>

** 한국과학기술정보연구원 지역혁신연구센터 지역AX연구팀장
(교신저자)

- ORCID: <https://orcid.org/0009-0006-5346-285X>

• Received: Jul. 13, 2026, Revised: Jul. 27, 2026, Accepted: Jul. 30, 2026

• Corresponding Author: Sanghun Cho

Div. of Data Analysis, Korea Institute of Science and Technology

Tel.: +82-2-3299-6098, Email: scho@kisti.re.kr

I. 서 론

지역의 AI 전환(AX, AI transformation)을 지원하는 정책 현장에서는 두 가지 유형의 진단 요구가 제기된다. 신규 기획된 AX 사업이 지향하는 대상과 수준에 대한 평가 요구와, 해당 사업을 수행할 지역 기업의 현재 역량 단계에 대한 진단 요구가 이에 해당한다. 전자는 계획을 대상으로 하고 후자는 실행을 대상으로 하지만 두 진단은 상호 독립적이지 않다. 계획의 실현 가능성과 지원 수단의 적절성을 판단하는 데 그것을 실행할 현장의 역량이 핵심 정보가 되기 때문이다[1][2]. 또한, 국내 실태조사에서 대기업과 중소기업, 수도권과 비수도권 간의 AI 활용 격차가 확인됨에 따라[3], 계획과 실행을 통합적으로 파악할 수 있는 진단에 대한 수요가 증가하고 있다. 이러한 맥락에서 지역의 기업 역량 분포와 해당 지역이 기획한 사업 목표를 동일 기준에서 검토하는 접근이 요구된다. 그러나 기존의 진단 도구는 이러한 두 진단을 단일 기준에서 처리하지 못하는 한계를 갖는다. 기업 진단 도구와 사업기획 평가 도구가 상이한 척도와 문항으로 구성되어 있어, 산출 결과를 동일 좌표에서 대조하는 것이 불가능하기 때문이다. 그 결과 계획이 실행과 정합하지 않는 경우, 예컨대 활용 역량이 앞선 지역이 공급 기술에 편중된 사업을 기획하는 경우에도 이를 식별하기 어렵다. 또한, 도구별로 프레임워크와 채점 논리가 중복 유지됨에 따라 유지보수 비용이 이중으로 발생하며, 기관이나 산업 클러스터 등 새로운 진단 대상이 요구될 경우 도구 전체의 재설계가 불가피하다. 이러한 문제의 근본 원인은 진단 대상의 차이를 시스템의 표층이 아닌 심층, 즉 측정 토대 수준에서 처리해 온 구조에 있다.

본 연구는 이 문제를 소프트웨어 및 계측 공학에서 정립된 하네스(Harness) 개념으로 접근한다. 하네스는 고정된 골격에 상황별 부품을 교체하여 다양한 대상을 동일 절차로 처리하는 구조를 의미한다(II장). 하네스 개념을 진단 시스템에 적용하면 측정 토대와 채점 방식은 고정하고 진단 대상과 성숙 단계만 교체하는 구조를 설계할 수 있다. 이러한 문제의식에서 본 연구는 두 가지 연구 질문을 설정한

다. 첫째(RQ1), 상호 연결된 복수의 AX 진단을 하나의 고정된 코어에서 처리하면서 대상별 차이는 교체형 부품으로 흡수하는 하네스를 어떻게 설계할 것인가? 둘째(RQ2), 그렇게 설계된 하네스는 실제 두 대상에 적용될 때 일관성과 분별력을 동시에 확보하는가?

이러한 접근이 요구되는 이유는 기존 대안의 구조적 한계에 있다. 대상별로 개별 도구를 개발하는 방식은 대상 적합성을 얻는 대신 결과의 비교 가능성을 상실하고, 모든 대상에 단일 문항 세트를 적용하는 범용 도구는 그 반대의 희생을 치른다. 두 속성이 동시에 요구되는 지역 AX 진단에서는 변화하는 요소(진단 대상과 성숙 단계)와 변하지 않아야 하는 요소(측정 토대·채점·집계)를 아키텍처 계층에서 분리하는 제3의 구조가 필요하다. 적응을 시스템의 재설계가 아니라 부품의 교체로 재정의한다는 점이 제안 구조가 기존 성숙도 모델 및 범용 진단 도구와 구별되는 지점이다.

본 연구의 기여는 진단의 측정 품질이 아니라 아키텍처에 있다. 첫째, 적응을 코어의 재설계가 아니라 부품의 교체로 재정의하고, 코어를 구성하는 세 가지 고정 요소와 두 가지 교체 요소를 정식화한다. 둘째, 진단의 공통 언어인 TIM 좌표계의 세 축을 설계 근거와 함께 정식화하고 그 구성을 표로 제시한다. 셋째, 기업 281개사와 사업기획 40건에 하네스를 적용하여 그 일관성과 분별력을 실증한다. 채점 엔진 자체의 측정학적 타당화는 별도의 영문 논문에서 다루며, 본 논문은 그 결과를 코어의 작동 근거로 요약 인용한다. 두 논문은 데이터의 일부를 공유하나, 연구 질문(하네스 설계 대 채점기 타당화), 분석 수준(시스템 구조 대 측정 척도), 기여 유형(설계 지식 대 측정 지식)이 상이하다.

II. 관련 연구

조직 및 사업의 역량을 단계별로 구분하여 위치시키는 성숙도 모델(Maturity model)은 정보시스템 관리 분야에서 지속적으로 활용되어 왔으며, 산업계의 AI 준비도 지수[4]에서 확인되듯 AI 영역에서도 성숙도 기반 접근이 확산되고 있다. 학술적으로는

Becker et al.[5]이 성숙도 모델 개발 절차의 체계성을 제시하였고, Pöppelbuß와 Röglinger[6]는 모델의 유용성을 결정하는 일반 설계 원칙을 정식화하였다. 그러나 이러한 모델의 대부분은 특정 진단 대상, 주로 대기업 조직을 전제로 설계된다는 한계를 갖는다. 이 경우 대상이 변경되면 문항과 준거를 재설계해야 하며, 이러한 대상 종속성은 서로 다른 모델의 결과를 동일 좌표에서 비교하는 것을 원천적으로 제약한다. 지역 정책이 요구하는 진단이 계획과 실패라는 인과 관계로 연결되어 있음을 고려하면 이 한계는 특히 중대하다. 흡수역량 이론[7]이 제시하듯 외부 기술의 가치는 이를 수용하고 활용하는 수요 측 역량에서 실현되며, 국내 실태조사[3]가 보여주는 규모·지역 간 활용 격차와 조직의 AI 역량을 다룬 실증 연구[8]도 기술 자산 못지않게 활용 역량이 성과를 좌우함을 보고한다. 따라서 계획의 기술 목표는 이를 흡수할 현장의 수요 역량과 함께 평가되어야 하나, 대상별로 분리된 도구는 이러한 통합적 평가를 지원하지 못한다. 여기서 제기되는 물음은, 진단 대상의 차이를 측정 토대가 아니라 그 상위의 교체 가능한 계층에서 처리할 수 있는가이다.

이러한 물음에 대한 단서는 소프트웨어 공학의 설계 개념에서 도출될 수 있다. 소프트웨어 시험에서 test harness는 피험 대상인 코드 단위를 교체하면서도 실행 및 검증의 골격은 고정하는 장치를 의미하며, 단위 시험 자동화의 표준 요소로 활용되고 있다[9]. 시스템 아키텍처 관점에서 안정적 핵심부와 교체 가능한 확장부를 분리하는 설계는 마이크로커널 및 플러그인 패턴으로 정식화되어 있다. Bass et al.[10]은 이러한 분리가 시스템의 수정가능성과 확장성을 높이는 핵심 전술임을 제시하였으며, Parnas[11]가 제안한 모듈 분해의 원리, 곧 변화하는 부분과 변화하지 않는 부분을 경계로 분리하라는 지침은 이 설계의 이론적 기반을 이룬다. 본 연구에서는 이러한 설계 원리를 진단 시스템에 적용하고자 한다. 즉, 측정 토대와 채점 엔진, 집계 규약은 고정된 코어로 유지하고, 진단 대상과 성숙 단계는 코어 외부의 교체 가능한 부품으로 분리하는 구조를 설계한다. 이를 통해 대상 종속성은 시스템의 심층에서 표층으로 이동하며, 코어가 고정됨에 따라

서로 다른 대상의 진단 결과가 동일 좌표에 축적될 수 있다. 코어 내부의 채점은 서로 다른 계열의 대형언어모델(LLM, Large Language Model)이 독립적으로 판정한 후 합의하는 방식을 따르며[12][13], 이는 단일 모델의 편향을 완화하기 위한 접근이며, LLM 판정의 신뢰성 자체가 검증을 요구한다는 지적[14]을 수용하여 본 연구는 채점 신뢰도를 직접 측정한다(V장). 주목할 점은 이 채점 방식이 진단 대상에 종속되지 않는다는 것이다. 기업 역량 진단과 사업기획 평가 모두 동일한 합의 엔진이 좌표별 수준을 판정하며, 대상의 차이는 채점 방식이 아니라 코어에 탑재되는 문항군으로만 표현된다.

III. 연구 방법

본 연구는 인공물의 설계와 평가를 통해 지식을 산출하는 설계과학연구(DSR, Design Science Research) 방법론을 채택하였다. DSR은 Hevner et al.[15]에 의해 정립되고 Peffers et al.[16]에 의해 실행 절차로 구체화된 방법론으로, 연구의 산출물은 문제를 해결하는 인공물과 이로부터 도출되는 설계 원칙으로 구성된다[17]. 본 연구의 인공물은 적용형 하네스에 해당한다. 표 1은 Peffers의 여섯 단계가 본 연구의 활동으로 어떻게 구체화되었는지를 제시한다.

표 1. 설계과학연구 단계와 본 연구의 활동
Table 1. DSR stages and corresponding research activities

DSR stage	Research activity	Sec.
(1) Problem identification	Target-dependent separation of linked diagnoses	I, II
(2) Objectives	Harness design requirements R1-R4	III
(3) Design and development	Fixed core with target/stage plug-ins	IV
(4) Demonstration	Application to firms and project plans	V
(5) Evaluation	Consistency, discriminability, quality	V
(6) Communication	Integrated reports and this paper	VI, VII

설계에 앞서 하네스가 충족해야 할 요구사항을 네 가지로 정식화하였다. 첫째, 코어를 구성하는 측정 토대와 채점 엔진, 집계 규약은 진단 대상 및 단계와 무관하게 고정되어야 한다(R1). 둘째, 대상별 차이는 코어가 아닌 교체형 문항군을 통해서만 표현되어야 한다(R2). 셋째, 성숙 단계에 따른 역량 중요도의 변화는 교체형 가중 프로파일에 의해 흡수되어야 한다(R3). 넷째, 새로운 진단 대상은 코어의 수정 없이 부품 추가만으로 수용 가능해야 한다(R4). 이 요구를 충족하는 하네스를 평가하기 위해 5개 권역과 9개 산업에 걸친 기업 281개사의 역량 진단 데이터와 지역 AX 사업계획 40건의 목표 평가 데이터를 사용하였다. 기업 데이터는 모두 비식별 처리하였으며, 두 대상의 척도가 상이하므로 프로파일 비교 시 표준화(z)하여 형태를 대조하였다.

IV. 적응형 하네스 아키텍처

그림 1은 하네스의 전체 구조를 제시한다. 그림 중앙의 굵은 테두리 영역이 모든 진단에서 고정되는 공통 코어이며, 코어는 세 계층으로 구성된다. 최상위 계층은 TIM 좌표계로서 진단 대상을 좌표

에 사상하고, 중간 계층은 다중 LLM 합의 엔진으로서 각 좌표의 수준을 판정하며, 최하위 계층은 집계 규약으로서 도메인 합성 점수와 단계 점수를 6대 4로 결합하여 최종 등급을 산출한다. 이 세 계층은 진단 대상과 무관하게 유지된다. 고정 요소와 교체 요소 사이의 이 경계가 곧 본 논문 전체 주장의 골격이다. 적응은 코어의 수정이 아니라 코어 양측의 두 소켓에 서로 다른 부품을 탑재함으로써 이루어진다. 좌측 소켓에는 진단 대상에 대응하는 문항군 어댑터가, 우측 소켓에는 성숙 단계에 대응하는 가중 프로파일이 탑재된다. 새로운 대상이 요구될 경우 코어를 유지한 채 어댑터를 추가하면 된다(R4). 여기서 명확히 할 점은, 본 연구의 적응이 모델의 재학습이나 파라미터추정을 수반하는 적응이 아니라는 것이다. 이는 고정된 골격에 부품을 교체하는 방식의 적응으로서, 보다 단순하고 예측 가능하다. 코어의 세 계층은 하나의 처리 흐름으로 연결된다. 진단이 시작되면 입력 문서(기업 제출 자료 또는 사업계획서)가 어댑터의 문항군과 함께 코어에 공급되고, TIM 계층이 대상을 기술·산업·성숙의 좌표에 사상한다.

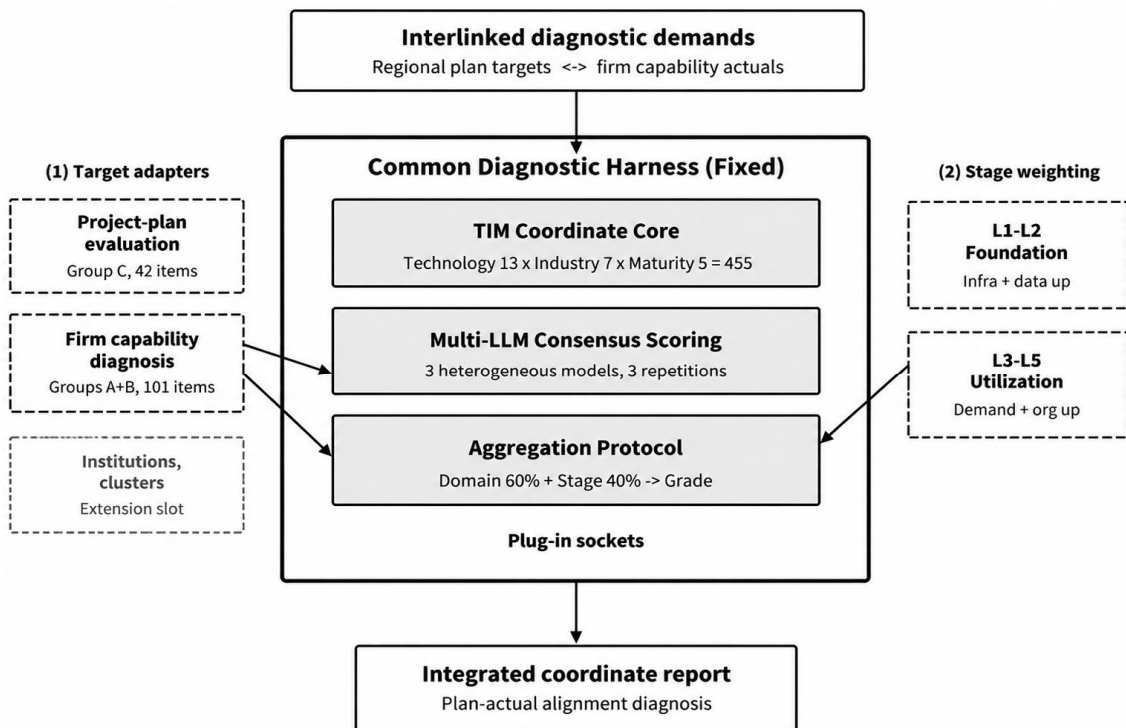


그림 1. AX 진단 하네스: 고정 코어와 교체형 플러그인 구조
Fig. 1. AX diagnostic harness: the fixed core and swappable plug-ins

이어 채점 엔진이 문항별로 좌표 수준을 판정하는데, 서로 다른 계열의 세 모델이 동일 입력에 대해 독립적으로 판정한 뒤 반복을 거쳐 합의 수준을 확정하며, 각 판정에는 입력 문서에서 발췌한 근거 문장의 인용이 요구된다. 이 근거 요구는 판정의 추적 가능성을 확보하는 장치로, V장에서 근거-수준 정합과 인용 충실도로 검증된다. 마지막으로 집계 계층이 도메인 합성 점수(60%)와 단계 점수(40%)를 결합하는데, 전자는 조직 역량의 현재 구성을, 후자는 발전 경로상의 위치를 반영하도록 배분한 설계 값으로, 집계 규약의 일부로서 코어에 고정된다. 이 구조는 III장의 설계 요구와 일대일로 대응된다. 세 계층의 고정은 R1에, 좌측 소켓의 문항군 어댑터는 R2에, 우측 소켓의 가중 프로파일은 R3에 대응하고 코어 수정 없는 어댑터 추가 규약이 R4를 구현한다. 이러한 추적성이 설계의 완결성을 보장한다. 판정된 문항 점수가 최종 등급으로 종합되는 집계 절차를 수식으로 정리하면 다음과 같다. 식 (1)과 같이 문항 i 의 합의 점수는 세 모델($M=3$)이 각 $R=3$ 회 반복 산출한 점수의 평균으로 정의된다.

$$s_i = \frac{1}{MR} \sum_{m=1}^M \sum_{r=1}^R s_i^{(m,r)} \quad (1)$$

도메인 d 의 점수는 식 (2)와 같이 해당 도메인 문항 집합 I_d 의 평균이며, 문항 수준에서는 등가 가중을 적용한다.

$$S_d = \frac{1}{|I_d|} \sum_{i \in I_d} s_i \quad (2)$$

최종 점수는 식 (3)과 같이 단계 조건 가중 $\alpha_d(L)$ 로 합성한 도메인 점수와 단계 점수 S_L 을 6대 4로 결합하여 산출된다.

$$F = 0.6 \cdot \sum_d \alpha_d(L) \cdot S_d + 0.4 \cdot S_L, \quad (3)$$

$$\sum_d \alpha_d(L) = 1$$

여기서 $\alpha_d(L)$ 은 성숙 단계 L 에 따른 도메인 중요

도의 이동을 표현하는 교체형 가중 프로파일로서, 초기 단계에서는 기반 도메인에, 성숙 단계에서는 수요·조직 도메인에 상대적으로 높은 값이 배분된다(그림 2는 그 설계 개념의 예시). 이 배분은 단계에 따라 요구 역량이 달라진다는 흡수역량 논의[7]에 근거를 두되, 구체 수치는 운영 설정으로 관리되는 설계값이며 배분 민감도의 정량 분석은 후속 연구에 다룬다. 코어의 측정 토대인 TIM 분류체계는 세 개의 좌표축으로 임의의 진단 대상을 위치시킨다. 어떤 AI 기술을 다루는가(Technology), 그 기술이 어떤 산업 맥락에서 활용되는가(Industry), 그 활용이 어느 발전 단계에 있는가(Maturity)라는 세 물음이 각각 하나의 축이 되며, 세 축의 조합은 $13 \times 7 \times 5 = 455$ 개의 분류 셀을 정의한다. 이는 측정이 완료된 공간이 아니라 진단 결과가 기록될 수 있는 좌표 체계를 뜻한다. 세 축은 서로 독립적인 정보를 담도록 설계되었다. 기술 축이 같아도 산업 맥락이 다르면 요구되는 데이터와 규제 환경이 다르고, 산업이 같아도 성숙 단계가 다르면 필요한 지원이 다르기 때문이다. 표 2는 세 축의 구성과 설계 근거를 요약한다. 이는 각 축의 구성이 임의의 선택이 아니라 외부 준거에 근거함을 보이기 위한 것이다.

표 2. TIM 좌표계의 세 축 구성과 설계 근거
Table 2. Three axes of the TIM taxonomy and their design rationale

Axis	Composition	Design rationale
Technology (T)	13 AI-technology subfields, refined from three families: discriminative/judgment AI, generative AI, and physical AI	Restructured for AX policy with reference to the OECD framework for classifying AI systems[18] and the technology categories of national AX policy
Industry (I)	Seven industry sectors (Table 3)	Consolidation of KSIC divisions into regional policy units
Maturity (M)	Five levels, L1-L5 (Table 4)	Established tradition of stage-based maturity models[5][6]

각 축의 입도(Granularity)는 정책 배분 단위와의 정합을 기준으로 결정하였다. 산업 축의 7개 부문(표 3)은 지역 주력산업 지정과 지원 프로그램이 운

용되는 단위에 대응하고, 기술 축의 13개 중분류는 AX 지원사업 공고에서 지원 대상 기술을 특정하는 분류의 상세화 수준에 대응하며, 성숙 축의 5개 수준(표 4)은 진단 결과가 연결될 개선 로드맵의 단계 수에 대응한다. 축을 더 잘게 나눌수록 표현력은 커지지만 455개 좌표에 대한 표본의 커버리지가 희박해져 좌표별 판정의 안정성이 떨어지므로, 이 입도는 표현력과 추정 안정성의 균형을 고려하여 설정한 것이며, 좌표 커버리지에 대한 정량 분석은 후속 과제다. 이 공통 좌표가 서로 다른 대상의 진단 결과를 동일 기준에서 비교하게 하는 장치이며, 하네스 코어의 측정 토대를 구성한다.

표 3. 산업 축(I)의 7개 부문

Table 3. Seven industry sectors of the I-axis

Sector	Coverage (example)
ICT and software	Telecommunications, software, platform services
Manufacturing	Machinery, electronics, and materials
Healthcare and bio	Medical devices, healthcare, bio
Environment and energy	Environment, energy, ESG-related
Agri-food and fisheries	Agriculture, food, fisheries
Construction	Construction and facility safety
Public and others	Public administration, logistics, others

표 4. 성숙 축(M)의 5개 수준

Table 4. Five maturity levels of the M-axis

Level	Name	Behavioral anchor (summary)
L1	Foundation	Recognizing AX needs; organizing basic resources
L2	Adoption	Trial application through pilot projects
L3	Diffusion	Extending validated use to multiple functions
L4	Institutionalization	AX integrated into processes and decisions
L5	Advancement	Leading AX-driven innovation; external diffusion

코어에 탑재되는 대상 어댑터는 진단 대상별 문항군으로 구성된다. 전체 문항은 143개로서 기업 역량 공통 진단용 A군 39문항과 심화 진단용 B군 62

문항, 사업기획 평가용 C군 42문항으로 구분된다(표 5). 사업기획 평가 시에는 계획의 목표수준과 충실도를 판정하는 C군이, 기업 역량 진단 시에는 조직의 현재 역량을 평가하는 A·B군이 코어에 공급된다. 여기서 주목할 점은 두 어댑터가 동일한 다중 LLM 합의 엔진을 공유한다는 것이다. 즉, 대상이 변경되어도 채점 방식은 유지되며, 입력 문서와 문항군만이 달라진다. 어떤 어댑터를 탑재하든 산출은 동일한 TIM 좌표와 7개 도메인(표 6)·5수준에 기록되므로, 계획과 실행을 동일 좌표에서 비교할 수 있다. 마지막으로 단계 가중 프로파일은 성숙 단계에 따른 역량 중요도의 변화를 흡수한다. 기반 구축 단계에서는 기술 인프라와 데이터 등의 토대 역량이 성과를 좌우하지만, 성숙 단계에서는 수요를 파악하고 조직을 운영하는 활용 역량이 보다 결정적이다[7].

표 5. 대상 어댑터를 구성하는 문항군

Table 5. Item groups constituting the target adapters

Group	Target	Items	Measurement focus
A	Firm capability (common)	39	Basic common capability
B	Firm capability (in-depth)	62	Domain-specific capability
C	Project plan	42	Target level and fidelity
Total	-	143	-

표 6. 산출이 기록되는 7개 도메인

Table 6. Seven domains for recorded outputs

Domain	Measured content
D2-1 Tech. infrastructure	Computing and network foundations
D2-2 Data	Data assets and governance
D2-3 Model and Algorithm	Model development and utilization
D2-4 HR and Organization	Talent, structure, and culture
D2-5 Process	AX integration into processes
D3 Supply ecosystem	External partnerships, supply chain
D4 Demand and Users	Market demand and user utilization

초기 단계(L1 - L2) 프로파일은 기반 도메인에 높은 가중을 부여하고, 성숙 단계(L3 - L5) 프로파일은

수요·활용과 인력·조직으로 가중을 이동시킨다(그림 2). 그림 2의 수치는 이러한 가중 이동의 설계 개념을 예시한 것이며, 실제 가중 프로파일은 운영 설정으로 관리되는 교체형 부품으로서 적용 맥락에 따라 조정된다. 따라서 이 프로파일의 교체는 코어의 집계 규약을 변경하지 않는다. 새로운 진단 대상의 수용 절차도 이 구조에서 자연히 도출된다. 예컨대 기관이나 산업 클러스터를 진단 대상으로 추가하려면, 해당 대상의 증거 문서 유형을 정의하고 이를 판정할 문항군을 작성하여 좌측 소켓의 규약(문항-좌표 대응, 근거 인용 형식)에 맞추어 탑재하면 된다. 코어의 어떤 계층도 수정되지 않으므로 기존 두 대상의 진단 결과와 새 대상의 결과가 자동으로 동일 좌표를 공유하며, 이것이 Parnas[11]의 정보 은닉이 진단 시스템에서 갖는 실질적 효용이다. 변화 가능성이 높은 결정(무엇을 진단하는가)은 어댑터에 격리되고, 변화해서는 안 되는 결정(어떻게 측정하고 비교하는가)은 코어에 보존된다.

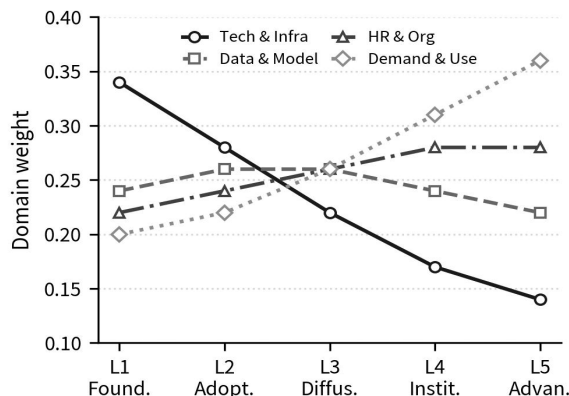


그림 2. 성숙 단계에 따라 교체되는 도메인 가중 프로파일(설계 예시)

Fig. 2. Stage-conditional domain weighting profiles(design illustration)

V. 하네스의 적용과 평가

하네스의 유효성은 두 가지 요건을 기준으로 평가된다. 첫 번째 요건은 일관성으로, 고정된 코어가 서로 다른 대상을 동일한 기준으로 평가하지 못하는 경우 공통 좌표의 의미가 형식에 그치게 된다. 두 번째 요건은 분별력이다. 코어가 모든 대상에 대해 유사한 결과만을 산출하는 경우, 이는 일관성이

아닌 변별력의 결여를 의미한다. 좋은 하네스는 두 대상을 같은 기준으로 읽으면서도 각 대상의 실제 차이를 놓치지 않아야 한다. 여기서 분별력은 판별 타당도의 통계적 확증이 아니라, 공통 좌표가 대상 간 차이를 표현해 내는지에 대한 확인을 뜻한다. 본장은 이 두 요건을 순서대로 검증하고(5.1 상당), 이어서 좌표계의 산업 축이 실질적 정보를 담는지와 코어 채점 자체의 측정 품질을 보조적으로 확인한다. 검증에 사용한 데이터는 III장에서 기술한 기업 281개사의 역량 진단과 사업기획 40건의 평가 결과 전체다. 적용 절차는 다음과 같다. 기업 진단에서는 기업별 제출 문서가 A·B군 어댑터와 함께 코어에 공급되어 7개 도메인 수준이 판정되었고, 사업기획 평가에서는 계획서가 C군 어댑터와 함께 동일한 코어에 공급되었다. 두 적용에서 코어의 어떤 요소도 수정되지 않았으므로, 이하의 결과는 곧 하나의 고정된 코어가 두 이질적 대상을 처리한 기록이다.

먼저 일관성에 대한 검토를 수행하였다. 두 대상은 문항군과 원칙도가 상이하므로(기업 1-5 수준, 사업기획 0-100점) 절대 점수의 직접 비교는 성립하지 않으며, 비교 가능한 것은 도메인 프로파일의 형태, 즉 일곱 개 도메인 중 상대적 고저의 패턴이다. 이에 따라 각 대상의 도메인 평균을 대상 내부에서 표준화(z)하여 척도 차이를 제거하고 형태를 대조하였다. 그림 3은 이렇게 표준화된 두 프로파일을 나란히 배치한 것으로, 막대가 0보다 위이면 그 대상의 평균보다 높은 도메인, 아래이면 낮은 도메인을 뜻한다. 두 프로파일의 상관은 $r=0.64$ 로, 일곱 개 도메인 쌍($n=7$)에 기반한 탐색적 지표이지만 기업과 사업기획 모두 인력·조직과 수요·활용이 높고 기술 인프라와 데이터가 낮은 유사한 상대 형태를 보였다. 서로 다른 문항군 어댑터를 탑재했음에도 하나의 코어가 유사한 상대 형태를 산출한 것이다. 이 형태 일치는 일관성 검토의 핵심 증거로서, 서로 다른 어댑터를 거친 두 산출이 동일 코어의 산출임을 보여 준다. 이 공통 형태 자체도 정책적 정보를 담는다. 계획과 실행이 공히 기술 인프라와 데이터, 곧 기반 영역에서 상대적 저점을 보인다는 것은 기반 조성형 지원을 우선 검토할 근거가 된다. 다만 이 상관은 일곱 개 도메인 값에 기반한 프로파일

형태의 유사성 지표이므로, 표본 상관처럼 통계적 유의성을 주장하는 값이 아니라 두 프로파일이 공유하는 구조의 크기를 기술하는 값으로 해석해야 한다.

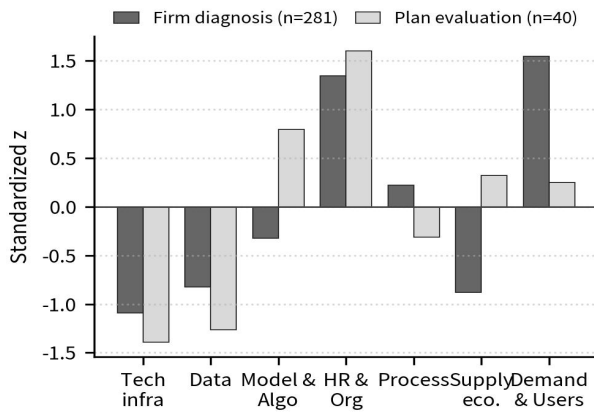


그림 3. 하나의 코어에서 두 진단 대상이 나타내는 도메인 프로파일

Fig. 3. Domain profiles of the two diagnostic targets on a single core

다음으로 분별력에 대한 검토를 수행하였다. 일관성이 두 대상이 공유하는 구조를 확인하는 것이라면, 분별력은 두 대상이 실제로 상이한 지점에서 상이하게 평가되는지를 확인하는 것이다. 이를 위해 공급 측 성숙(기술·데이터·모델 등 여섯 도메인의 평균)과 수요 측 성숙(수요·사용자 도메인)을 두 축으로 하는 평면에 전체 진단 대상을 배치하였다(그림 4). 이 평면은 분별력 검토의 요체로서 두 대상이 상이한 영역을 차지하는지를 직접 확인하게 한다. 평면의 대각선은 공급과 수요가 균형을 이루는 선으로, 대각선 위쪽은 수요가 공급을 앞서는 영역이다. 기업(원)은 공급 1.51·수요 2.09 부근, 곧 대각선의 뚜렷한 위쪽에 분포한다. 공급과 수요의 점수 차이를 표준화한 효과크기(대상별 차이점수의 평균을 그 표준편차로 나눈 값)는 $d_z = 1.06$ 으로서 통상 0.8 이상을 큰 효과로 해석하는 관례에 비추어 매우 큰 격차다. 이 수요 우위는 초기 단계 기업들이 활용 의지와 수요 인식은 갖추었으나 이를 뒷받침할 기술 기반이 미치지 못하는 실태를 보여 준다. 반면 사업기획(마름모)은 원점수(0 - 100)를 선형 변환(20 점=1수준)한 척도에서 공급 1.07·수요 1.10으로 평면의 좌하단, 대각선 부근에 밀집한다($d_z = -0.26$). 평가

대상 사업 다수가 기반 조성 단계의 지역 과제라는 구성적 특성이 반영된 위치이며, 계획 문서의 목표 설정 성향에 대한 해석은 후속 검증을 요한다. 요컨대 동일한 코어가 기업에서는 수요 우위라는 실태의 신호를, 사업기획에서는 저성숙 목표라는 계획의 신호를 각각 다른 좌표로 분별해 냈으며(표 7), 이것이 하네스가 요구하는 두 번째 요건의 충족이다.

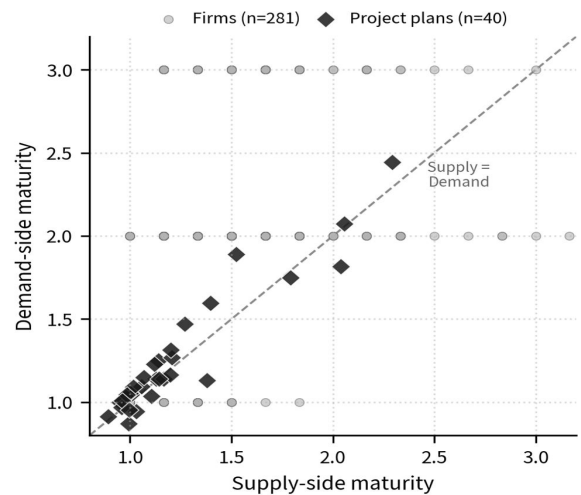


그림 4. 공급-수요 성숙도 평면상의 두 진단 대상

Fig. 4. Two diagnostic targets on the supply - demand maturity plane

표 7. 두 진단 대상의 좌표 통계 대조

Table 7. Coordinate statistics of the two diagnostic targets

Indicator	Firm diagnosis (n=281)	Plan evaluation (n=40)
Supply-side maturity (mean)	1.51 (1-5 scale)	1.07 (1-5 scale)
Demand-side maturity (mean)	2.09 (1-5 scale)	1.10 (1-5 scale)
Supply-demand gap d_z	-1.06	-0.26
Profile shape	HR/demand dominant (shared)	HR/demand dominant (shared)
Distribution	Demand-dominant	Low-maturity target range

또한, 좌표계의 산업 축이 명목적 구분에 그치지 않고 실질적인 분별 정보를 포함하는지에 대한 검토를 수행하였다. 기업 281개사를 산업 부문별로 구분하여 성숙 분포를 집계한 결과(표 8), 최상위 수준(L3)에 도달한 8개사는 ICT·소프트웨어 5개사와 의료·바이오 3개사로 한정되었으며, 제조·환경·에너지 등 나머지 부문은 모두 L1 - L2 수준에 분포하였다.

표 8. 산업별 성숙도 분포 (기업 n=281)

Table 8. Maturity distribution by industry (281 firms)

Industry	n	L1	L2	L3	Mean
ICT and SW	77	21	51	5	1.73
Healthcare and Bio	36	14	19	3	1.64
Agri-food and Fisheries	11	5	6	0	1.58
Environment and Energy	61	31	30	0	1.56
Manufacturing	50	27	23	0	1.51
Others	46	27	19	0	1.40-1.76 (range)

같은 기술이라도 어느 산업에 놓이느냐에 따라 성숙의 진행 속도가 다르다는 것이며, 이는 산업을 독립 축으로 설계한 선택(표 2)의 실증적 근거가 된다. 특히 표본이 50개사에 이르는 제조 부문에서 L3가 전무하다는 사실은 표본 크기만으로는 설명되지 않아 부문 특성으로 해석할 여지가 있으며(기업 규모·업력 등 교란 요인의 통제 후속 과제다), 데이터 축적 환경과 디지털 전환 선행도가 유리한 ICT·의료 부문에서 상위 수준이 먼저 출현하는 양상과 정합적이다. 정책적으로는 상위 수준 기업이 존재하는 부문과 하단에 집중된 부문에 서로 다른 지원 유형이 필요함을 시사한다. 코어의 채점 엔진이 신뢰할 만한 좌표 판정을 산출하는지 확인하였다. 채점에는 서로 다른 계열의 세 모델인 claude-sonnet-4-6(Anthropic), gpt-4o-mini(OpenAI), Llama-3.3-70B-Instruct-Turbo(Meta)를 동일한 문항·프롬프트·설정으로 각 3회 반복 실행하였으며, 접근 경로와 프롬프트, 합의 규칙을 포함한 실행 명세는 재현성 확보를 위해 표준화된 절차서로 문서화되어 관리된다. 급내상관계수(ICC, Intraclass Correlation Coefficient)는 서로 다른 평가자가 같은 대상에 얼마나 일치된 점수를 주는가를 나타내는 지표로, 여기서는 세 LLM을 세 평가자로 두고 산출하였다. 괄호 안 표기는 산출 조건을 뜻한다. ICC(2,1)은 평가자 한 명의 점수를 그대로 사용할 때의 신뢰도이고 ICC(2,k)는 k명(여기서는 3개 모델)의 평균 점수를 사용할 때의 신뢰도이다. 사업기획 40건에 대해 단일 모델의 신뢰도는 ICC(2,1)=0.455에 그쳤으나 세 모델 평균 점수의 신뢰도는 ICC(2,k)=0.715로 크

게 상회하였다. 점수 변동을 분해한 결과 대상 간 분산이 46%, 모델 간 관점 차이가 21%, 잔차가 33%로 나타났다. 이는 단일 모델에 의존할 경우 판정의 약 5분의 1이 모델 선택이라는 우연적 요인에 좌우됨을 의미한다. 내용타당도는 문항들이 측정하려는 영역을 적절히 대표하는 정도를 뜻한다. 검증을 위한 전문가 패널은 연구계(AI·표준·국방), 산업계(AI 솔루션·클라우드·금융), 학계, 지역 산업진흥기관에 걸친 전문가 11인으로 구성하여 공급·수요·정책의 관점이 함께 반영되도록 하였다. 각 전문가는 143개 문항 전체에 대해 측정 목표와의 관련성을 4점 척도로 독립 평정하였고, 문항별 관련성 지수(I-CVI)를 산출한 뒤 그 평균으로 척도 수준 지수(S-CVI/Ave)를 종합하였다. 그 문항 평균인 S-CVI/Ave는 0.877로 통상 기준(0.80)을 충족하였다. 아울러 채점이 실제로 근거에 기반했는지를 보기 위해, 시스템이 제시한 채점 근거만 보고 수준을 역추론했을 때 실제 부여 수준과 일치한 비율(근거-수준 정합)은 77.1%였고, 근거로 제시된 인용문이 실제 입력 문서에 존재하는 문장인지 확인한 인용 충실도(근거의 실재성 지표로, 근거가 결론을 지지하는지와는 구분된다)는 100%였다(표 9). 이 지표들은 코어 채점을 신뢰할 수 있는 조건과 유보해야 할 조건을 함께 명시함으로써 산출 해석의 경계를 제공한다.

표 9. 코어 채점의 측정 품질 요약

Table 9. Measurement quality of the core scoring engine

Metric	Value	Interpretation
ICC(2,1), single model	0.455	Moderate
ICC(2,k), 3-model average	0.715	Improved by consensus
Variance (between-target/model/residual)	46/21/33%	Single-model reliance inadequate
Content validity, S-CVI/Ave	0.877	Meets criterion (>=0.80)
Rationale-level concordance	77.1%	Response-process evidence
Citation fidelity	100%	Traceable evidence
Expert-criterion correlation	0.07	Not established (support only)

한편 시스템 점수를 알지 못하는 상태에서 독립적으로 평정한(차폐된) 전문가 점수와 상관은 $r=0.07$ 에 그쳤다. 이 결과는 네 가지 요인을 함께 고려하여 해석할 필요가 있다. 첫째, 전문가 평정이 수행된 소형 과제 26건의 소표본에서 산출된 값이라는 표집의 한계가 있다. 둘째, 전문가는 현장 지식을, 시스템은 제출 문서만을 정보원으로 삼는 정보 기반의 비대칭이 존재한다. 셋째, 평가 대상 다수가 저성숙 구간에 밀집하여 점수 분산이 제한되었으며, 분산 제한은 상관계수를 하향 편위시키는 요인으로 알려져 있다. 넷째, 준거 점수는 계획당 차폐된 전문가 2인의 평정에 기반하는데, 소수 평정자의 평균만으로는 준거 자체의 측정 오차가 충분히 제거되지 않으며 준거의 오차는 상관을 감쇠시키는 요인으로 알려져 있다. 이 요인들은 낮은 상관이 곧 시스템의 무효를 의미하지 않음을 시사하지만, 준거 타당도가 확립되지 않았다는 사실 자체는 남는다. 이에 본 논문은 채점 사용의 타당성을 Kane[19]의 논증 기반 틀에 따라 다음과 같이 구성한다. 채점 추론(판정이 규칙과 근거에 따라 이루어졌는가)은 근거-수준 정합 77.1%와 인용 충실도 100%가, 일반화 추론(판정이 모델과 반복에 걸쳐 안정적인가)은 세 모델×3회 반복의 분산 구조와 $ICC(2,k)=0.715$ 가, 내용 추론(문항이 측정 영역을 대표하는가)은 $S-CVI/Ave=0.877$ 이 각각 뒷받침한다. 반면 외삽 추론(점수가 전문가가 판단하는 실제 수준과 일치하는가)은 현 단계에서 $r=0.07$ 로 지지되지 않으며, 이것이 산출의 사용을 서열화나 선별이 아닌 의사결정 지원으로 한정하는 직접적 이유다. 곧 제안 시스템의 유효성 주장은 준거 일치에 대한 것이 아니라, 지지되는 세 추론의 범위 안에서 정의된 사용(비교 가능한 좌표 정보의 제공)에 대한 것이며, 이 범위의 판단은 본 논문에 제시된 증거로 충분하다. 외삽 추론의 확립에는 전문가에게 시스템과 동일한 문서만 제공하는 정보 접근 통제, 평정 표본의 확대, 평정자 수 확대와 평가자 간 신뢰도의 명시적 산출이 요구되며, 이 요건은 진행 중인 후속 검증의 설계에 반영되어 있다. 평가 점수를 공적 결정에 사용하려면 그 해석과 사용의 정당화가 선행되어야 하므로[19], 준거 타당도가 확립되기 전까지

하네스의 산출은 의사결정을 대체하기보다 지원하는 역할에 한정된다. 끝으로 본 장의 검증 구조를 정리해 둔다. 아키텍처 수준에서는 고정 코어가 두 대상에서 일관성과 분별력을 동시에 확보하는지를 확인하였고, 채점 수준에서는 앞서의 논증 구조에 따라 채점·일반화·내용 추론의 근거를 제시하였다. 두 수준의 증거가 결합되어 의사결정 지원이라는 정의된 사용 범위를 뒷받침한다.

VI. 논 의

두 대상에 대한 하네스 적용 결과는 세 가지 설계 원칙으로 정리되며, 각 원칙은 선행 설계 이론과 연결된다. 첫째, 코어의 불변성이다. 측정 토대와 채점 엔진, 집계 규약이 대상 및 단계와 무관하게 고정됨으로써 서로 다른 진단의 결과가 동일 좌표에서 비교될 수 있다. 이는 Pöppelbuß와 Röglinger[6]가 성숙도 모델의 유용성 조건으로 제시한 비교 가능성 요구를, 개별 모델의 설계 지침이 아니라 아키텍처 계층의 불변 조건으로 확장한 것이다. 둘째, 적응의 부품 교체 한정이다. 대상 어댑터와 단계 프로파일의 교체만으로 적응이 완결됨으로써 코어의 수정 없이 새로운 대상의 수용이 가능해진다. 이는 변화 가능성이 높은 결정을 모듈 내부에 은닉하도록 하는 Parnas[11]의 정보 은닉 원리를 진단 시스템에 적용한 것에 해당한다. 진단 대상은 변화 가능성이 가장 높은 설계 결정이므로 코어 밖의 어댑터에 격리되어야 하며, 실제로 본 연구의 두 어댑터는 코어의 어떤 요소도 수정하지 않고 탑재되었다. 셋째, 모든 산출을 공통 좌표에 기록한다. TIM 좌표계에 결과를 기록하면 계획과 실행과 같이 상호 연결된 진단을 통합적으로 검토할 수 있다.

제안 설계의 위치는 현행 실무의 두 가지 대안과의 비교를 통해 명확해진다. 첫 번째 대안은 대상별 개별 도구를 유지하는 방식으로, 각 대상에 최적화된 문항을 활용할 수 있으나 본 연구에서 지적한 비교 불가능성과 중복 유지보수의 문제를 해소하지 못한다. 두 번째 대안은 모든 대상에 단일 문항 세트를 일괄 적용하는 범용 도구 방식이다. 이 경우 비교 가능성은 확보되지만, 계획 문서와 조직 실행

라는 상이한 증거에 같은 문항을 적용하는 데서 오는 대상 부적합이 발생한다. 하네스는 이 두 대안의 절충이 아니라 제3의 구조다. 문항은 대상별로 최적화되(어댑터), 좌표와 채점과 집계는 공통으로 묶어(코어), 대상 적합성과 비교 가능성을 동시에 달성한다(표 10). 이 대비는 제안 구조가 두 대안의 절충이 아니라 제3의 구조임을 보여 준다. 실증 결과가 보인 일관성($r=0.64$)과 분별력(그림 4의 상이한 분포)의 동시 확보는 이 구조적 선택의 직접적 결과다.

표 10. 대안 구조와 제안 하네스의 비교

Table 10. Comparison of alternative structures and the proposed harness

Structure	Target fit	Comparability	Maintenance
Target-specific tools	High	Low	Duplicated logic
Single universal items	Low	High	Single item set
Generic plug-in design	High	Optional	Core reuse
Proposed harness	High (adapters)	Ensured by fixed core	Core reuse

제안 구조의 특징은 세 가지로 요약된다. 첫째, 서로 다른 대상에 대한 진단 결과가 동일한 좌표 위에서 비교될 수 있다. 둘째, 새로운 진단 대상이 요구될 때 코어의 수정 없이 어댑터의 추가만으로 확장이 가능하다. 셋째, 채점 모델이 교체되더라도 그 영향이 엔진 계층에 국한되어 문항과 좌표 체계는 그대로 유지된다. 다만 이러한 특징은 각각의 전제 위에서 성립하며, 각 전제에 대한 실증은 후속 연구로 확장된다.

하네스 구조는 채점 기술의 변화에 대해서도 함의를 갖는다. LLM은 세대 교체가 빈번한 기술이며, 특정 모델에 결합된 진단 도구는 모델이 바뀔 때마다 전면 재검증에 직면한다. 본 아키텍처에서 채점 엔진은 코어의 한 계층으로 캡슐화되어 있고 문항·좌표·집계와 분리되어 있으므로, 모델 교체의 영향은 엔진 계층에 국한된다. 교체 시 기록된 설정으로 재검증을 거치면 문항과 좌표 체계는 그대로 유지되며, 이는 급변하는 기반 기술 위에서 진단의 연속성을 확보하는 실무적 장치가 된다. 정책 활용의 관

점에서 하네스가 제공하는 새로운 기능은 계획과 실행을 동일 좌표에서 대조하는 정합 진단이다. 예컨대 광역 단위의 AX 지원 사업을 설계할 때, 관내 기업들의 도메인 프로파일(그림 3)을 기준선으로 두고 신규 사업기획의 목표 좌표를 대응시키면, 수요 활용이 이미 앞선 지역이 공급 기술에 편중된 사업을 과도하게 기획하는지, 혹은 기반이 취약한 산업군에 성숙 단계를 생략한 목표가 설정되어 있는지를 심사 단계에서 식별할 수 있다. 또한 표 8과 같은 산업별 성숙 분포는 지원 프로그램의 차등 설계, 곧 상위 수준 기업이 존재하는 산업에는 선도·후발 연계형을, 하단에 집중된 산업에는 기반 조성형을 배치하는 근거 자료가 되며, 이는 지역 AI 확산을 국정 의제로 둔 정책 방향[20]과도 부합한다. 대상별로 분리된 도구로는 수행하기 어려웠던 이러한 연계 분석은, 상호 연결된 두 진단을 하나의 코어로 처리한 하네스 설계의 직접적 결과다.

본 연구의 한계와 그 함의는 다음과 같다. 첫째, 하네스의 확장성(R4)은 설계상의 주장이며, 본 연구는 기업과 사업기획 두 어댑터의 적용으로 이를 예시했을 뿐 기관이나 산업 클러스터와 같은 제3의 어댑터의 탑재는 검증하지 못하였다. 확장성의 완전한 입증은 새 어댑터가 코어 수정 없이 탑재되는 사례의 축적을 요구한다. 둘째, 두 대상의 척도 차이로 프로파일 비교가 표준화에 의존하였다. 형태의 일관성은 확인되었으나 절대 수준의 직접 비교, 예컨대 특정 사업기획의 목표가 관내 기업 평균보다 몇 수준 높은지의 판정에는 두 척도를 연결하는 정규화 연구가 선행되어야 한다. 셋째, 코어 채점의 전문가 준거 타당도가 미확립 상태이므로, 하네스의 산출을 배제적 결정의 단독 근거로 사용하는 것은 현 단계에서 적절하지 않다. 이 한계는 아키텍처의 결합이 아니라 채점 엔진의 성숙도 문제이며, 엔진이 코어의 독립 계층으로 분리되어 있으므로 준거 타당화의 진전이 곧 하네스 전체의 신뢰 상승으로 이어지는 구조다.

VII. 결 론

본 연구에서는 상호 연결된 AX 진단을 단일 고

정 코어에서 처리하는 적응형 하네스 아키텍처를 제안하고, 기업 281개사와 사업기획 40건에 대한 적용을 통해 적용 가능성과 탐색적 근거를 제시하였다. 두 연구 질문에 대한 결과는 다음과 같이 정리된다. 첫째(RQ1), 진단 하네스는 TIM 좌표계·다중 LLM 합의 엔진·집계 규약의 3계층 고정 코어와, 대상 어댑터·단계 가중 프로파일의 2중 교체형 플러그인으로 설계될 수 있으며, 적응은 코어의 재설계가 아니라 부품의 교체로 정의된다. 둘째(RQ2), 이렇게 설계된 하네스는 두 대상의 도메인 프로파일에서 유사한 상대 형태($r=0.64$)를 산출하는 동시에 기업의 수요 우위($dz=-1.06$)와 사업기획의 저성숙 목표($dz=-0.26$)를 서로 다른 좌표로 표현하였다. 채점 코어의 평가자 간 신뢰도($ICC(2,k)=0.715$)와 내용타당도($S-CVI/Ave=0.877$)는 코어 작동의 기초 요건을 지지하나, 전문가 준거와의 관계는 확립되지 않아 산출의 사용은 의사결정 지원에 한정된다.

이 과정에서 코어 불변, 부품 적응, 공통 좌표라는 세 설계 원칙을 도출하였다. 이 원칙들은 본 시스템에 국한되지 않는다. 진단 대상이 다양해 결과의 비교 가능성이 요구되는 영역, 예컨대 디지털 전환·탄소중립·연구보안과 같이 계획 평가와 실행 진단이 함께 요구되는 정책 영역이라면, 측정 토대를 코어로 고정하고 대상 차이를 어댑터로 격리하는 동일한 구조가 적용될 수 있다. 그런 의미에서 본 연구의 기여는 특정 진단 도구가 아니라, 연계된 진단들을 하나의 좌표에서 처리하는 설계 지식이다.

후속 연구는 다음과 같은 방향으로 확장해 나가자 한다. 첫째, 기관·산업 클러스터 어댑터를 실제로 탑재하여 확장성 요구(R4)를 사례로 검증한다. 둘째, 두 대상의 척도를 연결하는 정규화 연구를 수행하여 형태 비교를 넘어 절대 수준의 비교를 지원한다. 셋째, 정보 접근을 통제된 전문가 평정과 확대 표본으로 준거를 재설계하여 코어 채점의 준거 타당도를 확립한다. 넷째, 계획과 실행을 동일 좌표에서 대조하는 정합 진단을 광역 단위 사업 기획·심사 현장에 파일럿으로 적용하여 실무 효용을 검증한다. 다섯째, LLM 세대 교체 시 엔진 계층만을 재검증하는 표준 절차를 정립하여 기반 기술 변화 위에서 진단의 연속성을 확보해 나갈 계획이다.

References

- [1] OECD, "AI Adoption by Small and Medium-sized Enterprises", OECD Publishing, Paris, 2025.
- [2] H. S. Sactra, "Generative AI: Here to Stay, but for Good?", *Technology in Society*, Vol. 75, Art. no. 102372, Nov. 2023. <https://doi.org/10.1016/j.techsoc.2023.102372>.
- [3] Korea Chamber of Commerce and Industry and Korea Institute for Industrial Economics and Trade, "Survey on AI Technology Utilization of Korean Firms", Aug. 2024. https://www.korcham.net/nCham/Service/Economy/appl/KcciReportDetail.asp?SEQ_NO_C010=20120938915&CHAM_CD=B001. [accessed: Jul. 01, 2026]
- [4] Cisco, "Cisco AI Readiness Index 2025", Cisco Systems, 2025. https://www.cisco.com/c/m/en_us/solutions/ai/readiness-index.html.
- [5] J. Becker, R. Knackstedt, and J. Poeppelbuss, "Developing Maturity Models for IT Management", *Business & Information Systems Engineering*, Vol. 1, No. 3, pp. 213-222, Jun. 2009. <https://doi.org/10.1007/s12599-009-0044-5>.
- [6] J. Poeppelbuss and M. Roeglenger, "What Makes a Useful Maturity Model? A Framework of General Design Principles for Maturity Models", *Proc. 19th European Conf. on Information Systems (ECIS 2011)*, Helsinki, Finland, Paper 28, Jun. 2011.
- [7] W. M. Cohen and D. A. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation", *Administrative Science Quarterly*, Vol. 35, No. 1, pp. 128-152, Mar. 1990. <https://doi.org/10.2307/2393553>.
- [8] P. Mikalef and M. Gupta, "Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance", *Information & Management*, Vol. 58, No. 3, Art. no. 103434, Apr. 2021. <https://doi.org/10.1016/j.im.2021.103434>.

- [9] P. Hamill, "Unit Test Frameworks", O'Reilly Media, Sebastopol, CA, 2004. <https://www.oreilly.com/library/view/unit-test-frameworks/0596006896/>.
- [10] L. Bass, P. Clements, and R. Kazman, "Software Architecture in Practice", 3rd ed., Addison-Wesley, 2012.
- [11] D. L. Parnas, "On the Criteria to Be Used in Decomposing Systems into Modules", Communications of the ACM, Vol. 15, No. 12, pp. 1053-1058, Dec. 1972. <https://doi.org/10.1145/361598.361623>.
- [12] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. P. Xing, H. Zhang, J. E. Gonzalez, and I. Stoica, "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena", Advances in Neural Information Processing Systems, Vol. 36, pp. 46595-46623, Dec. 2023. <https://doi.org/10.48550/arXiv.2306.05685>.
- [13] P. Verga, S. Hofstatter, S. Althammer, Y. Su, A. Piktus, A. Arkhangorodsky, M. Xu, N. White, and P. Lewis, "Replacing Judges with Juries: Evaluating LLM Generations with a Panel of Diverse Models", arXiv:2404.18796, Apr. 2024. <https://doi.org/10.48550/arXiv.2404.18796>.
- [14] K. Schroeder and Z. Wood-Doughty, "Can You Trust LLM Judgments? Reliability of LLM-as-a-Judge", arXiv:2412.12509, Dec. 2024.
- [15] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research", MIS Quarterly, Vol. 28, No. 1, pp. 75-105, Mar. 2004. <https://doi.org/10.2307/25148625>.
- [16] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research", Journal of Management Information Systems, Vol. 24, No. 3, pp. 45-77, Dec. 2007. <https://doi.org/10.2753/MIS0742-1222240302>.
- [17] S. Gregor and A. R. Hevner, "Positioning and Presenting Design Science Research for Maximum Impact", MIS Quarterly, Vol. 37, No. 2, pp. 337-355, Jun. 2013. <https://doi.org/10.25300/MISQ/2013/37.2.01>.
- [18] OECD, "OECD Framework for the Classification of AI Systems", OECD Digital Economy Papers, No. 323, OECD Publishing, Paris, Feb. 2022. <https://doi.org/10.1787/cb6d9eca-en>.
- [19] M. T. Kane, "Validating the Interpretations and Uses of Test Scores", Journal of Educational Measurement, Vol. 50, No. 1, pp. 1-73, Mar. 2013. <https://doi.org/10.1111/jedm.12000>.
- [20] Government of the Republic of Korea, "123 National Policy Tasks of the Lee Jae-myung Administration", Sep. 2025. <https://www.korea.kr/govVision/>. [accessed: Jul. 01, 2026]

저자소개

최 광 훈 (Kwang hun Choi)



2020년 2월 : 한양대학교
기술경영학과(기술경영학석사)
2022년 8월 : 한양대학교
기술경영학과(기술경영학박사)
2023년 7월 ~ 현재 :
한국과학기술정보연구원
지역혁신연구센터장

관심분야 : 인공지능 전환, 과학
기술 데이터 분석, 온톨로지 시스템, HAI 시스템

조 상 훈 (Sanghun Cho)



2017년 7월 : 서울대학교
나노융합학과(공학박사)
2018년 3월 ~ 2022년 7월 :
기술보증기금 차장
2022년 7월 ~ 현재 :
한국과학기술정보연구원
지역AX연구팀장

관심분야 : 인공지능 전환, 중소기업 AX 지원 시스템,
과학기술 데이터 분석, 지능형 아카이브 시스템