서울대학교 SEOUL NATIONAL UNIVERSITY

AI 학습데이터 전환과 통계적 추론의 공진화

공공데이터를 문서와 수치 자료에서 코드, API, 프롬프트, 에이전트 실행 이력까지 포함하는 국가 학습자산으로 재정의하기

고길곤

서울대학교 행정대학원 교수
정보화본부 본부장

pubdata.snu.ac.kr/lecture/policydata/

서울대학교 SEOUL NATIONAL UNIVERSITY

AI 시대, 국가데이터의 의미가 바뀌고 있다

국가데이터연구원의 데이터, 코드, 업무 절차는 AI가 정책문제를 이해하고 실행할 수 있는 형태인가?

기존 공공데이터 정책

- 개방 건수와 활용 실적 중심
- 민간 서비스 개발과 행정 효율화 지원
- 데이터셋 단위의 관리와 공개

AI 시대의 다음 질문

- 복지, 고용, 교육, 지역 데이터가 정책문제 단위로 연결되어 있는가
- 데이터의 생성 근거, 수집 방식, 품질 한계가 메타데이터로 남아 있는가
- 지표와 분석 결과를 만든 코드와 파이프라인을 재현할 수 있는가
- AI 에이전트의 판단 로그와 사람의 승인 이력이 감사 가능한가

서울대학교 SEOUL NATIONAL UNIVERSITY

지능형 정책 AI 플랫폼: Agent들의 결합

질문형성 Agent
정책문제를 하위 질문, 이해관계, 필요한 근거로 분해

통계수집 Agent
공공데이터, 행정자료, 통계표, 문서를 탐색하고 수집

통계분석 Agent
코드와 모델을 실행해 지표, 추세, 이상값, 불확실성을 분석

정책형성 분석 Agent
대안별 효과, 비용, 위험, 가치 충돌과 집행 가능성을 비교

정책제언 Agent
정책대안, 근거표, 설명자료, 보고서 초안과 후속 과제를 제안

공유 정책지능 레이어

- 국가데이터·공공데이터
- 문서·법령·민원
- 분석 코드·API
- 프롬프트·실행 로그

질문 기반 정책 | **근거 기반 정책** | **학습하는 정책체계**

서울대학교 SEOUL NATIONAL UNIVERSITY

개방에서 학습으로: 공공데이터의 다음 단계

3

공공데이터의 미래 가치는 개방 건수보다 AI가 학습할 수 있는 품질, 맥락, 연결성에서 결정된다.

품질 | 맥락 | 연결성 | 재현성 | 책임성

공개 중심	활용 중심	학습 중심
AI의 강력한 검색과 데이터 결합 기능을 통해 공개 데이터의 발견 가능성과 연결의 질이 깊어진다.	다부처, 다분야, 다가치 정책문제에 대해 질문 기반, 근거 기반 정책을 가능하게 하고 정책 소통의 질을 높인다.	정책학습을 통해 과거의 시행착오와 반복적 실패위험을 데이터로 식별하고 줄인다.

서울대학교 SEOUL NATIONAL UNIVERSITY

AI 학습데이터는 문서와 수치 데이터에 그치지 않는다

01

문서

법령, 지침, 보고서, 회의록, 민원, 정책자료. 제도와 판단 맥락을 담은 언어 데이터.

02

양적 데이터

통계, 행정 DB, 공공데이터셋, 센서 및 이용 데이터. 정책 상태와 변화를 담은 구조화 데이터.

03

코드

분석 코드, 처리 스크립트, 자동화 로직, API, 시뮬레이션 모델. 실행 가능한 지식.

04

에이전트

프롬프트, 도구 사용 방식, 판단 규칙, 워크플로, 실행 로그. 업무 수행 방식의 기록.

AI는 정보생산 도구를 넘어, 검색하고 결합하고 실행하는 업무수행 Agent로 확장되고 있다.

정책자료 조사

법령, 선행연구, 해외사례를 검색하고 쟁점별 근거표와 요약안을 생성

행정데이터 분석

공공데이터를 결합하고 분석 코드를 실행해 지표, 그래프, 이상값을 점검

민원·보고 업무

민원 유형을 분류하고 관련 규정 확인, 답변 초안, 후속 조치 목록을 작성

서울대학교 SEOUL NATIONAL UNIVERSITY

코드도 학습데이터다: 실행 가능한 정책지식

코드가 담고 있는 것

- 데이터가 어떤 기준으로 정제되고 결합되는가
- 정책 지표가 어떤 계산식으로 산출되는가
- 분석 결과가 어떤 가정과 필터를 거쳐 생성되는가
- 반복 업무가 어떤 순서로 자동화되는가

간략한 예시

- GitHub는 코드, 수정 이력, 이슈, 검토 과정을 함께 축적해 재현 가능한 업무 지식으로 만든다.
- Codex의 skill은 반복 업무 절차와 도구 사용 방식을 명시해 AI가 같은 방식으로 다시 수행하게 한다.
- 따라서 코드는 결과물이 아니라 정책 분석과 행정업무의 절차를 학습시키는 데이터가 된다.

서울대학교 SEOUL NATIONAL UNIVERSITY

에이전트도 학습데이터다: 업무 수행 방식의 학습

에이전트가 남기는 기록

- 프롬프트와 시스템 지시
- 에이전트가 던진 질문과 하위 과제 분해 방식
- 도구 호출과 API 사용 이력
- 중간 판단, 선택지, 오류 수정 과정
- 사람-AI 협업과 승인 절차

서울대 사례: 프롬프트 정보의 데이터화

- 질문, 프롬프트, 응답, 피드백을 단순 이용 흔적이 아니라 AI 서비스 개선 데이터로 관리
- 좋은 질문과 실패한 질문을 교육, 연구, 행정 업무별 프롬프트 자산으로 전환
- 민감정보와 보안 기준을 적용해 책임 있는 활용과 감사 가능성을 확보
- 핵심은 답변이 아니라 문제를 묻는 방식 자체를 조직학습 데이터로 삼는 것

Agent 질문 로그
반복적으로 요청되는 정책질문, 통계분석, 자료 탐색, 오류 수정 요구를 축적

질문 군집화
저출산, 지역격차, 고용, 복지, 안전 등 자주 등장하는 연구질문 묶음 발견

연구수요 분석
어떤 데이터가 부족한지, 어떤 분석기법과 지표가 반복적으로 필요한지 파악

연구의제 환류
국가데이터연구원의 데이터 구축, 방법론 연구, 교육, 품질관리 과제로 연결

국가데이터연구원은 Agent가 남긴 질문과 실행 기록을 분석해, “무엇이 자주 연구질문의 대상이 되는가”를 데이터로 파악할 수 있다.

서울대학교 SEOUL NATIONAL UNIVERSITY

공공데이터를 학습데이터로 전환하기 위한 조건

01

문제 중심성

어떤 국가적 난제를 풀기 위한 데이터인지 명확해야 한다.

정책수요 · 의제 · 성과지표

02

맥락성

데이터가 어떤 제도와 현장, 시점에서 생성되었는지 설명되어야 한다.

법령 · 절차 · 지역 · 시점

03

연결성

기관과 분야를 넘어 같은 문제 단위로 결합될 수 있어야 한다.

표준 · 식별자 · 메타데이터

04

재현성

분석과 판단을 다시 검증할 수 있도록 코드와 절차가 남아야 한다.

코드 · 파이프라인 · API

05

실행성

AI 업무 수행 과정이 기록되고 다음 판단을 개선해야 한다.

프롬프트 · 도구 호출 · 로그

학습데이터 전환은 데이터셋 정리가 아니라, 질문·맥락·연결·재현·실행이 이어지는 정책지식 인프라를 만드는 일이다.

서울대학교 SEOUL NATIONAL UNIVERSITY

AI 통계추론은 국가데이터를 정책지능으로 바꾸는 장치다

국가데이터

통계표, 행정자료, 공공데이터, 문서와 법령이 정책문제의 원재료가 된다.

통계 | 행정 DB | 문서 | 메타데이터

→

AI 통계추론

질문 형성, 자료 탐색, 모형 제안, 코드 실행, 진단과 해석을 하나의 절차로 묶는다.

질문 | 가정 | 코드 | 검증 로그

→

정책지능

반복 가능한 근거 생산과 학습을 통해 정책 판단의 질과 설명 가능성을 높인다.

근거 | 시나리오 | 대안 비교 | 소통

AI 통계추론의 핵심은 통계를 대신하는 것이 아니라, 국가데이터가 정책 질문에 답하는 과정을 재현 가능하고 학습 가능한 형태로 바꾸는 것이다.

서울대학교 SEOUL NATIONAL UNIVERSITY

국가데이터 기반 AI 통계추론 워크플로

01
질문정의
정책문제를 분석 가능한 질문과 성과지표로 전환

02
자료탐색
국가데이터, 공공데이터, 행정자료, 문서, 메타데이터 연결

03
모형제안
추정, 예측, 인과추론, 시뮬레이션 방법 후보 비교

04
코드실행
Python, R, SQL, API를 통해 재현 가능한 계산 수행

05
검증·보고
불확실성, 한계, 출처, 해석을 감사 가능한 기록으로 남김

국가데이터처의 과제는 더 많은 통계표를 제공하는 것에서, 정책 질문이 검증 가능한 분석 절차로 이어지도록 만드는 것으로 확장된다.

서울대학교 SEOUL NATIONAL UNIVERSITY

DGP와 식별: 데이터만으로 정책 효과를 확정할 수 없다

$D_n \sim P_M,\ P_{M_1}(D)=P_{M_2}(D) \Rightarrow \delta(D_n)$는 M_1과 M_2를 구분할 수 없음

동일한 관측분포를 만드는 두 생성 과정이 있으면, 데이터만 보는 어떤 절차도 둘을 식별할 수 없다는 뜻이다.

- D_n: n개의 관측값으로 이루어진 표본 데이터
- M, M_1, M_2: 데이터를 만들어낸 후보 생성 과정 또는 모형
- P_M: 모형 M이 만들어내는 관측 데이터의 확률분포
- $\delta(D_n)$: 표본을 보고 어느 모형인지 판단하는 통계적 절차

DGP의 의미

- 관측된 통계는 현실 전체가 아니라 생성 과정의 일부 표본
- 정책, 제도, 지역, 집행 방식, 개인 선택이 함께 데이터를 만듦
- 통계추론은 숫자를 맞히는 것이 아니라 생성 과정을 좁혀 가는 작업

식별의 의미

- 동일한 관측 패턴이 여러 정책 메커니즘으로 설명될 수 있음
- 선택편의, 누락변수, 역인과, 측정오차가 정책 효과 해석을 흔듦
- 무작위 배정, 자연실험, 도구변수, 패널 비교가 식별을 보강

AI는 식별 문제를 없애지 않는다. 다만 가능한 DGP와 식별 가정을 빠뜨리지 않도록 점검하는 보조자가 될 수 있다.

서울대학교 SEOUL NATIONAL UNIVERSITY

가정은 약점이 아니라 데이터와 정책 현실을 잇는 다리다

$$Y_i = m(X_i) + \varepsilon_i, \quad E[\varepsilon_i \mid X_i] = 0$$

회귀계수의 정책적 해석은 오차항이 설명변수와 체계적으로 연결되어 있지 않다는 가정 위에서만 가능하다.

가정	정책분석에서의 역할	AI가 보조할 수 있는 점검
비교가능성	정책 대상과 비교집단이 어떤 조건에서 비교 가능한지 규정	집단 차이, 제도 차이, 사전추세, 누락된 교란요인 후보 탐색
측정 타당성	지표가 정책 목표와 현장 변화를 제대로 대표하는지 판단	지표 정의, 조사 방식, 분류 변경, 행정자료 생성 규칙 확인
모형 적합성	선형성, 독립성, 분포, 시간 의존성 같은 분석 조건을 명시	잔차, 이상치, 군집, 계절성, 비선형 패턴의 진단 코드 제안

좋은 AI 통계추론은 가정을 숨기는 것이 아니라, 어떤 가정 위에서 어떤 결론이 가능한지 공개하는 방식이어야 한다.

서울대학교 SEOUL NATIONAL UNIVERSITY

추정과 검정은 코드 실행보다 진단과 해석이 중요하다

$$\hat{\theta} \pm z_{\alpha/2} \cdot SE(\hat{\theta}), \quad p = Pr(T \geq t_{obs} \mid H_0)$$

추정치는 불확실성과 함께 제시되어야 하며, 검정은 귀무가설 아래 관측값이 얼마나 이례적인지 묻는 절차다.

STEP 1

추정

평균 차이, 회귀계수, 효과크기, 예측값을 산출하되 데이터 정의와 표본 범위를 함께 기록한다.

STEP 2

검정

p-value만이 아니라 신뢰구간, 사전 가설, 다중비교, 표본 크기의 영향을 함께 해석한다.

STEP 3

진단

잔차, 이상치, 결측, 분포 변화, 민감도 분석으로 결과가 얼마나 강건한지 확인한다.

STEP 4

보고

결론, 불확실성, 한계, 대안 해석, 후속 데이터 필요성을 함께 제시한다.

AI통계추론 덱의 PAL 관점 반영: 논리와 해석은 LLM이 보조하되, 계산은 Python·R·SQL 같은 도구가 실행하고 로그로 남겨야 한다.

서 울 대 학 교
SEOUL NATIONAL UNIVERSITY

인과추론에서 AI는 식별전략을 제안하지만 식별을 보장하지 않는다

$$ATE = E[Y_i(1) - Y_i(0)], \quad \{Y_i(1), Y_i(0)\} \perp T_i \mid X_i$$

정책효과는 관측되지 않는 반사실과의 차이이며, 조건부 비교가능성 같은 식별 가정 없이는 직접 관측되지 않는다.

AI가 도울 수 있는 일

- 정책 개입, 결과변수, 매개변수, 교란변수 후보를 구조화
- DAG, 반사실 질문, 대안적 설명을 빠르게 생성
- 자연실험, DID, RDD, IV, 매칭 등 후보 설계를 비교
- 민감도 분석과 위약 검정 코드 초안을 작성

사람이 책임져야 할 일

- 제도와 현장을 이해해 비교 가능성의 근거를 판단
- 측정 오류와 집행 과정의 변화를 확인
- 식별 가정이 깨지는 사례를 반례로 검토
- 정책 결론의 윤리적·정치적 함의를 설명

AI는 인과효과를 증명하는 기계가 아니라, 식별전략과 반례를 더 체계적으로 검토하게 만드는 방법론적 동료다.

서울대학교 SEOUL NATIONAL UNIVERSITY

인과효과 측정 기법은 “어떤 비교가 정당한가”를 묻는다

기법	핵심 아이디어	AI가 보조할 수 있는 부분
RCT·A/B test	무작위 배정으로 처치집단과 비교집단의 평균적 비교가능성 확보	실험 설계, 표본수 검토, 이질효과 탐색, 실험 로그 품질 점검
DID·event study	정책 전후 변화와 비처치집단 변화를 비교해 공통추세 가정 활용	사전추세 진단, 정책 시행시점 정리, 위약 검정 코드 작성
RDD·IV	경계값 또는 도구변수가 만드는 준실험적 변동을 이용	제도 규칙 탐색, 조작 가능성 점검, 1단계 강도와 국소효과 해석 보조
Matching·IPW	관측 공변량이 유사한 집단을 만들거나 가중치를 부여해 비교	공변량 후보 탐색, 균형성 진단, overlap 부족 사례 탐지

서울대학교 SEOUL NATIONAL UNIVERSITY

Double Machine Learning: 예측 AI를 인과추정의 보조 도구로

$$Y = \theta D + g(X) + \varepsilon, \quad D = m(X) + v, \quad E[v\varepsilon] = 0$$

AI는 g(X)와 m(X) 같은 nuisance function을 유연하게 추정하고, 관심 효과 θ는 orthogonal score로 편향을 줄여 추정한다.

왜 필요한가

- 정책 대상자의 배경변수가 많고 비선형 관계가 강한 경우가 많음
- 단순 회귀는 공변량 조정이 부족하거나 모형 지정 오류에 취약
- 머신러닝은 예측에는 강하지만, 그대로 쓰면 인과효과 추정 편향 가능

핵심 절차

- 처치모형과 결과모형을 별도로 학습
- cross-fitting으로 과적합 편향을 완화
- 잔차화된 처치와 결과의 관계에서 정책효과 추정
- 표준오차와 신뢰구간으로 불확실성 제시

참고: Chernozhukov et al.의 Double/Debiased Machine Learning. 핵심은 “AI가 효과를 직접 말하는 것”이 아니라 “교란 조정을 더 잘하게 하는 것”이다.

서울대학교 SEOUL NATIONAL UNIVERSITY

이질적 처치효과는 “누구에게 효과가 큰가”를 묻는다

$$\tau(x) = E[Y(1) - Y(0) \mid X = x]$$

평균처치효과 ATE가 전체 평균을 묻는다면, CATE는 지역, 연령, 소득, 제도 환경에 따라 효과가 어떻게 달라지는지 묻는다.

METHOD 1

Causal Tree

효과 차이가 큰 하위집단을 나무 구조로 분할해 설명 가능성을 높인다.

METHOD 2

Causal Forest

여러 나무를 결합해 개별 특성별 처치효과와 불확실성을 추정한다.

METHOD 3

Uplift Modeling

개입했을 때 추가로 변화할 가능성이 큰 대상을 찾아 정책 타겟팅에 활용한다.

METHOD 4

Policy Learning

제약된 예산 안에서 누구에게 어떤 정책을 우선 적용할지 규칙을 학습한다.

참고: causal tree, generalized random forest, causal forest 계열 방법론은 “평균 효과”에서 “분배된 효과”로 정책분석의 질문을 확장한다.

서울대학교 SEOUL NATIONAL UNIVERSITY

AI와 인과추론의 결합은 분석 파이프라인 전체를 바꾼다

01
DAG 초안
문헌, 법령, 현장 지식에서 변수와 경로 후보 생성

02
교란 점검
누락변수, 선택편의, 측정오차 후보를 질문 목록화

03
기법 추천
RCT, DID, RDD, IV, DML, causal forest의 적합성 비교

04
코드 감사
전처리, 추정, 표준오차, 민감도 분석 코드 검토

05
반례 탐색
가정이 깨지는 사례와 대안 해석을 체계적으로 제시

LLM은 인과관계의 최종 판정자가 아니라, 연구설계의 누락을 줄이고 검증 질문을 더 촘촘하게 만드는 분석 동료로 써야 한다.

최근 LLM 기반 causal discovery 연구는 가능성과 한계를 함께 보여준다. 보수적인 활용 원칙은 “결정은 방법론과 연구설계가, AI는 탐색과 검증 보조가 맡는다”이다.

서울대학교 SEOUL NATIONAL UNIVERSITY

AI는 국가데이터의 정책 활용을 네 가지 방식으로 보조한다

ROLE 1

질문 설계자

정책문제를 분석 단위, 변수, 비교집단, 기간, 성과지표로 분해한다.

ROLE 2

자료 연결자

통계표, 행정자료, 법령, 보고서, 민원, 현장 지식을 연결한다.

ROLE 3

계산 실행자

코드와 도구를 호출해 추정, 검정, 예측, 시각화를 재현 가능하게 수행한다.

ROLE 4

검증 보조자

가정, 편향, 결측, 이상치, 해석 오류와 불확실성을 점검한다.

AI통계추론의 핵심 메시지: LLM은 계산기가 아니라 질문, 도구, 데이터 출처, 해석 맥락을 연결하는 정책 분석 코디네이터다.

AI 시대의 통계품질은 불확실성을 표현하는 능력까지 포함한다

AI가 위험해지는 지점

- 통계적 한계를 자연어의 확신으로 덮어버림
- 상관관계를 인과관계처럼 설명
- 표본, 측정, 결측, 분류 기준의 차이를 생략
- 최신 데이터와 출처를 확인하지 않고 답변 생성

국가데이터 관리의 기준

- 효과크기, 신뢰구간, 예측구간, 민감도 분석을 함께 제시
- 데이터 정의와 변경 이력을 답변 근거에 포함
- 모형 가정과 대안 해석을 나란히 기록
- 사람의 검토와 승인 이력을 감사 로그로 보존

AI가 더 그럴듯하게 말할수록, 국가데이터 체계는 표본, 측정, 모형, 해석의 불확실성을 더 명시적으로 말해야 한다.

에이전틱 통계분석: LLM + 도구 + 데이터 출처

기존 LLM만으로 어려운 일

- 대용량 통계표와 행정자료 전체를 한 번에 이해
- 정확한 수치 계산과 재현 가능한 분석 수행
- 최신 통계, 원자료, 메타데이터 출처를 자동 검증
- 분석 코드, 프롬프트, 판단 이력을 체계적으로 남김

국가데이터 에이전트의 설계

- 필요한 데이터만 API와 데이터베이스에서 선택 조회
- Python, R, SQL 등 검증 가능한 도구로 계산 실행
- 차트, 표, 보고서 초안을 생성하고 근거를 연결
- 오류, 수정, 승인 과정을 다음 학습데이터로 축적

국가데이터 플랫폼은 LLM이 모든 데이터를 기억하게 만드는 곳이 아니라, 필요한 데이터와 도구를 정확히 호출하게 만드는 실행 환경이어야 한다.

서울대학교 SEOUL NATIONAL UNIVERSITY

국가데이터연구원의 미래는 AI 통계지능 인프라다

전통적 역할	AI 시대의 확장 역할	국가데이터 체계와의 연결
표본설계, 조사방법, 지표 개발, 통계품질 진단	AI 공식통계 방법론, 합성자료, 행정자료 결합, 통계 코드 검증	AI가 개입한 통계 생산과 해석을 공적으로 신뢰할 수 있는 기준 마련
통계작성기관의 방법론 지원	공공 통계 에이전트의 벤치마크, 오류분류, 품질 인증 체계 구축	부처와 지자체가 활용하는 AI 분석 도구의 최소 품질기준 제시
사후 품질관리와 교육	프롬프트·질문 은행, 분석 코드 저장소, AI 통계분석 샌드박스 운영	실패한 질문과 검증된 분석 절차를 조직학습 자산으로 전환

미래의 국가데이터연구원은 통계 생산의 후방 연구조직을 넘어, 국가데이터가 AI와 결합될 때 무엇을 신뢰할 수 있는 근거로 인정할지 설계하는 기관이 되어야 한다.

서울대학교 SEOUL NATIONAL UNIVERSITY

정책역량 이론과 AI 시대의 쟁점

정책역량의 차원	핵심 의미	AI 시대의 쟁점
분석역량	문제를 정의하고 근거를 수집·해석해 정책대안을 설계하는 능력	AI가 제시한 근거의 출처, 편향, 불확실성을 검증할 수 있는가
운영역량	정책을 집행하고 조직, 절차, 자원을 안정적으로 작동시키는 능력	코드, API, 에이전트가 행정절차 안에서 책임 있게 작동하는가
정치·조정역량	다부처, 다분야, 다가치 갈등을 조정하고 사회적 수용성을 확보하는 능력	AI 기반 정책판단을 시민, 국회, 부처, 현장과 설명 가능하게 소통하는가

AI는 정책역량을 대체하지 않는다. 다만 정책역량의 중심을 “답을 찾는 능력”에서 “좋은 질문을 만들고 근거와 실행을 검증하는 능력”으로 이동시킨다.

서울대학교 SEOUL NATIONAL UNIVERSITY

정책지능이란 무엇인가

정책지능의 의미

- 정책문제를 정확히 정의하고 질문으로 전환하는 능력
- 데이터, 문서, 현장 지식에서 근거를 찾고 해석하는 능력
- 대안의 효과, 비용, 위험, 가치 충돌을 비교하는 능력
- 집행 결과를 학습해 다음 정책 판단을 개선하는 능력
- 근거와 판단을 이해관계자에게 설명하고 소통하는 능력

AI의 역할

- 분산된 국가데이터와 문서를 빠르게 검색하고 연결
- 복잡한 정책문제를 하위 질문과 분석 과제로 분해
- 코드와 모델을 실행해 근거, 시나리오, 위험 신호를 생성
- 에이전트 로그를 축적해 정책 실패와 반복 업무를 학습
- 국회, 시민, 현장, 부처별 설명 자료를 맞춤형으로 작성

정책지능은 더 많은 정보를 갖는 것이 아니라, 더 좋은 질문과 검증 가능한 근거로 정책 판단의 질을 높이는 능력이다.

서울대학교 SEOUL NATIONAL UNIVERSITY

AI 시대 정책역량의 재정의

기존 정책역량

- 의제 설정
- 정책 분석
- 이해관계 조정
- 집행과 평가

AI 시대 추가 역량

- 정책문제를 데이터 문제로 번역
- 국가데이터를 연결하고 해석
- 분석 코드와 파이프라인을 재현
- 에이전트의 판단과 실행 절차를 검증
- 공공 AI 활용의 위험과 책임을 관리

서울대학교 SEOUL NATIONAL UNIVERSITY

국가 데이터처에서 지능정보처로

데이터처의 전통적 역할	지능정보처로 발전할 때의 역할
데이터 개방 관리자	국가 학습데이터와 정책지능 인프라 설계자
부처 간 데이터 중개자	다부처, 다분야 정책문제를 해석하는 지능정보 조정자
데이터 포털 운영자	공공 AI가 검색, 결합, 분석, 설명할 수 있는 지식 플랫폼 운영자
데이터셋 품질 점검자	AI가 생산한 정보의 출처, 품질, 편향, 재현성을 관리하는 신뢰 기관
단기 활용 실적 관리자	데이터, 코드, 프롬프트, 에이전트 로그를 축적해 정책학습을 이끄는 기관

데이터처의 다음 단계는 데이터를 관리하는 기관을 넘어, 국가의 정책판단을 지능화하는 지능정보처로 발전하는 것이다.

서울대학교 SEOUL NATIONAL UNIVERSITY

데이터 거버넌스의 관리 대상도 확장되어야 한다

기존 관리 대상

- 데이터셋
- 메타데이터
- 품질과 표준
- 개인정보와 보안

새로운 관리 대상

- 분석 코드와 모델링 파이프라인
- API와 자동화 로직
- 프롬프트와 도구 사용 정책
- 에이전트 실행 로그와 승인 이력

핵심은 모든 것을 공개하는 것이 아니라, 정책 판단의 재현성과 책임성을 확보할 수 있는 관리 체계를 갖추는 것이다.

서울대학교 SEOUL NATIONAL UNIVERSITY

AI가 학습할 수 있는 국가데이터 체계

> **AI를 잘 쓰는 국가는 데이터를 많이 가진 국가가 아니라, 공공데이터와 실행 지식을 정책지능으로 전환할 수 있는 국가다.**

기록

행정자료, 공공데이터, 문서, 코드, 프롬프트, 에이전트 로그를 함께 남긴다.

Data | Code | Prompt | Log

→

학습자산

질문, 가정, 분석 절차, 검증 결과를 재사용 가능한 정책지식으로 구조화한다.

질문은행 | 분석절차 | 품질기준

→

정책지능

국가의 정책 경험을 AI가 학습 가능한 공공 인프라로 재구성한다.

근거생산 | 정책학습 | 책임성