K-문샷 Initiative
Enterprise Private AI
2026

새결AI
백서

GIST 기획혁신본부
GIST ITRC 블록체인지능융합센터
리버밴스(주) LiberVance

새결AI
Enterprise Private AI

기관의 지식자산과 업무 맥락을 보존하는 Private AI
Private AI for shaping new knowledge
contexts from your organization's research flow

백서 · 사업 제안서
(White Paper · Project Proposal)

제안자: GIST 이흥노 교수

2026. 6. 16.

Executive Summary

새결AI(SaeGyeol AI) 는 “기관의 지식자산과 업무 맥락을 보존하는 Enterprise Private AI”이다. AI for Science 시대에 연구는 거대한 반복 탐색 문제가 되었고, 이를 가속하는 가장 빠른 길은 Big Tech 초거대 AI를 쓰는 것이지만, 그 대가로 실패 실험 로그·공정 조건·미공개 후보·특허 가능 아이디어가 외부로 흘러간다. 새결AI는 단일 모델이 아니라, 무엇을 외부로 보내고 무엇을 기관 안에 남길지를 판단하는 **Private AI Control Plane**이며, 보호(Privacy)와 탐색(Search / Optimization)을 하나의 폐루프에서 동시에 최적화한다.

새결AI의 핵심은 두 개의 Control Plane을 하나로 통합한 **이중 제어면(dual control plane)** 구조이다.

- **프라이버시 제어면(Privacy Control Plane).** 입력·파일·도구 호출을 민감도로 분류하고, 외부 글로벌 AI · 국내 독자 모델 · 로컬 GPU · 신뢰실행환경(TEE) · 사람 승인 중 실행 경로를 선택한다 [1].
- **탐색·최적화 제어면(Search & Optimization Control Plane).** 과학 발견을 트리 탐색 문제로 보고, Tree search / pruning / Bayesian optimization으로 다음 후보 · 다음 실험 · 다음 추론 경로를 선택한다 [28,29].
- **통합 지점.** 다음 후보 x와 실행 경로 route를 함께 선택한다. 정보이득과 프라이버시 누출을 공동 최적화하는 **프라이버시-제약 폐루프 실험설계**가 본 제안의 신규성이다.

1차 적용 사례는 **새결 Private AI Solar MissionLoop**(페로브스카이트/탠덤 태양전지)이며, 이 구조는 K-문샷 12대 미션 중 GIST가 직접 정합하는 8개 미션의 공통 후방 AI 도구층으로 확장된다 [56]. 한 줄 결론은 다음과 같다 — 새결AI는 “외부 AI를 쓰지 말자”는 시스템이 아니라, 연구 생산성과 데이터 주권을 동시에 확보하는 데이터 주권형 AI Scientist 인프라이다.

1. 문제 제기: AI4Sci 시대, 연구 방식이 바뀌고 있다

AI4Sci Korea 2026의 투고 독려문은 두 가지 질문을 던진다 — “혹시 나는 아직 예전 방식으로 연구하고 있는 건 아닐까?”, “AI를 연구에 써야 한다는 건 알겠는데, 실제로 과학 연구는 어떻게 바뀌고 있을까?” 이는 학술행사 홍보 문구가 아니라, 연구의 작업 방식 자체가 바뀌고 있다는 신호다.

- **연구 자동화가 “지능형 과학 발견”으로 확장되고 있다.** 최근 연구들은 AI Agent가 문헌 조사 · 가설 생성 · 실험 설계 · 실험 실행 · 결과 분석 · 논문 작성 · 동료평가까지 연구 전주기의 일부 또는 전부를 수행함을 보였다 [2,4,5]. Robin은 문헌·데이터분석 Agent를 연결해 생물학 실험의 가설–실험–해석을 반복했고 [2], The AI Scientist는 아이디어 생성부터 자동 리뷰까지 종단간 자동화를 실증했다 [4].
- **변화의 본질은 “연구가 거대한 반복 탐색 문제가 되었다”는 점이다.** 연구자는 더 많은 후보를 더 빠르게 검토하고, 실패를 더 빨리 학습하며, 근거 있는 다음 실험을 선택해야 한다. 따라서 AI4Sci 시대의 핵심 인프라는 단순 챗봇이 아니라 지식·실험·업무·문서·검토를 연결하는 연구 운영체계이다.
- **단, 2026년 현재의 자율성은 도메인이 좁고 도구가 정형화된 환경에 한정된다.** 예컨대 The AI Scientist가 생성한 논문은 수용률 70%의 워크숍 1차 심사를 통과한 수준이며 [4], 후속 연구도 미성숙한 아이디어·환각 등을 주요 실패모드로 보고한다 [6]. 핵심 병목은 모델의 언어 능력이 아니라 도구 연동의 견고성 · 실험 환경 표준화 · 출처 검증 · 안전한 폐루프 운영이다 [59]. 새결AI는 바로 이 운영 계층을 겨냥한다.

2. 핵심 9단 Narrative

AI Agent가 자율적으로 실험을 수행하며 과학적 발견의 주역이 될 수 있음이 속속 증명되면서(1단계), 선도 사례들은 Big Tech의 초거대 AI를 빌려 AI 과학자·엔지니어를 빠르게 만들어 냈지만, 그 과정에서 개인정보·업무 노하우·지식재산이 외부로 흡수될 위험과(2–3단계) 서비스 중단·가격 인상·약관 변경이라는 통제 불가능한 의존이 함께 드러났다(4단계). 다행히 오픈소스·Agentic AI 생태계가 빠르게 성장하면서, SOTA 모델을 로컬 GPU에 올려 기관이 직접 통제하는 Private AI를 구축할 길이 열렸다(5–6단계). 한편 과학 발견 자체는 찾고·설계하고·실험하고·쓰고·검토받고·고치는 거대한 반복 탐색 문제이며, 이런 문제는 트리 탐색·가지치기·베이지안 최적화로 효과적으로 풀 수 있다(7–8단계). 따라서 이 두 줄기를 합치면 기관의 데이터 주권을 지키면서 연구를 가속하는 Private AI Agent 시스템을 과학적 발견에 투입할 수 있으며(9단계), 새결AI는 바로 이 결론을 구현한 데이터 주권형 AI Scientist 인프라다. 아래 표는 그 9단계를 요약한 것이며, 본 백서의 각 절은 이 골격을 따른다.

단계	핵심 문장	제안서 의미
1	AI Agent는 자율 실험을 수행하고 Scientific Discovery에서 주요 역할을 할 수 있음이 증명되고 있다.	AI4Sci는 연구 보조가 아니라 연구 방식의 전환이다.
2	선도 사례들은 Big Tech 초거대 AI를 활용해 AI Scientist와 AI Engineer를 만든다.	빠른 실증은 가능하지만 외부 AI 의존이 커진다.
3	Big Tech AI는 서비스 과정에서 개인정보·업무노하우·지식재산을 흡수할 수 있다.	기관 데이터 주권과 IP 보호가 핵심 리스크가 된다.
4	SOTA AI는 편리하지만 서비스 중단·가격 인상·약관 변경 위험이 있다.	국가·기관 핵심 인프라를 외부 정책에만 맡길 수 없다.
5	Open Source 모델과 Agentic AI는 다수 개발자가 성능 개선에 참여·공유한다.	개방형 생태계로 기술 자립성과 확장성을 확보한다.
6	SOTA 모델을 local GPU에 구축·운용하면 Private AI를 만들 수 있다.	로컬 GPU·온프레미스·국내 소버린 AI가 전략 자산이 된다.
7	과학 발견은 찾고·설계하고·실험하고·쓰고·리뷰받고·고치는 거대한 반복 탐색 문제이다.	새결AI는 연구 전 과정을 loop로 관리해야 한다.
8	거대한 반복 탐색 문제는 Tree search·pruning·Bayesian optimization으로 잘 풀 수 있다.	후보·실험·문서·검토 경로를 선택하는 최적화 엔진이 필요하다.
9	Private AI Agent 시스템을 Scientific discovery에 활용할 수 있다.	새결AI는 데이터 주권형 AI Scientist 인프라가 된다.

3. 왜 Enterprise Private AI가 필요한가

3.1 외부 초거대 AI의 장점과 한계

외부 초거대 AI는 최신 모델을 즉시 쓸 수 있고 긴 문서 요약 · 코드 작성 · 복잡한 추론에 강하다. 그래서 초기 AI Scientist 시스템은 외부 AI를 적극 활용한다 [4,9]. 문제는 연구기관의 데이터가 일반 소비자 데이터와 다르다는 점이다.

보호 대상	예시	외부 유출 시 위험
개인정보	연구 참여자·학생·환자·평가자 정보	법적 책임, 신뢰 훼손
연구 데이터	실패 실험 로그, 측정 곡선, 원시 데이터	재현 가능한 IP 유출
공정 조건	온도·농도·장비 설정·공정 window	영업비밀 유출
특허 가능 아이디어	후보 조성, 신규 구조, 실험계획	선출원·권리 귀속 분쟁 [25]
업무 맥락	누가·무엇을·왜 결정했는지의 로그	기관 지식자산 손실

3.2 외부 AI 의존의 운영 리스크

외부 AI는 성능뿐 아니라 운영 정책에 의존한다. 서비스 중단 · 가격 인상 · 약관 변경 · 데이터 반출 · 공급망 종속은 기관이 통제하기 어려운 변수다. 새결AI는 (i) 로컬·국내·외부 모델의 다중 라우팅, (ii) 비용 미터링과 월 비용 cap, (iii) 민감 데이터의 원칙적 내부 처리, (iv) 마스킹·추상화, (v) MCP / 도구 표준으로 이 리스크에 대응한다 [17,18,19].

4. 새결AI의 정의

4.1 한 문장 정의

새결AI는 조직의 연구 흐름에서 새로운 지식 맥락을 빚어내는 Private AI이며, 기관의 지식자산과 업무 맥락을 보존하면서 외부 AI · 국내 독자 파운데이션 모델 · 로컬 GPU 모델 · 내부 데이터 · 업무 도구 · 연구 장비를 안전하게 연결하는 **Enterprise Private AI Control Plane**이다.

4.2 “Private AI is a routing decision, not a model choice”

공개 논문 요약은 외부 고성능 AI가 더 잘한다. 반대로 내부 실패 로그 · 특허 초안 · 공정 조건은 외부로 보내면 안 된다. 따라서 Private AI의 본질은 로컬 모델 하나를 설치하는 일이 아니라, 어떤 데이터와 업무 맥락을 어디로 보낼지 결정하는 라우팅이다 [1,17,18].

5. 새결AI 아키텍처

새결AI는 기존 시스템을 대체하지 않고 연결·제어하는 **Control Plane**이다. 모든 데이터와 작업은 Private Orchestration Layer를 반드시 경유하며, 이 계층이 "무엇을 어디로 보낼지"를 판단한다.

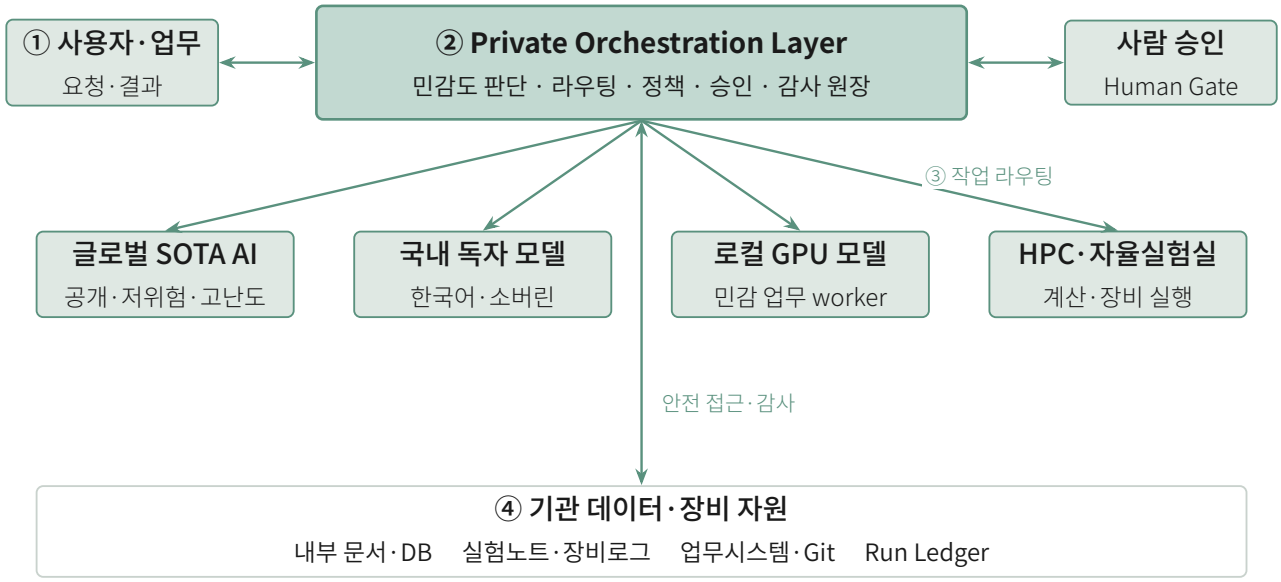

Figure 1. 새결AI Private AI Control Plane. 사용자 요청은 Orchestration Layer로 들어오고(① → ②), 제어면이 작업의 민감도·난도·비용을 판단해 적절한 모델·계산 자원으로 라우팅하며(② → ③), 기관 데이터·장비 자원은 제어면을 통해서만 안전하게 접근된다(② ↔ ④). **모델과 데이터가 직접 연결되는 경로는 존재하지 않는다** — 외부로 나가는 모든 정보는 제어면의 마스킹·정책을 거친다.

5.1 4계층 구조와 각 계층의 역할

새결AI는 위에서 아래로 네 계층으로 구성된다. ① **사용자·업무 계층**은 연구·행정·기업 사용자의 자연어 요청을 받는 접점이다. ② **Orchestration 계층**은 시스템의 두뇌로, 입력의 민감도를 판단하고 실행 경로를 정하며 도구를 호출하고 모든 처리를 기록한다. ③ **모델·계산 계층**은 제어면이 호출하는 "실행기"로, 작업 성격에 따라 글로벌 SOTA·국내 독자 모델·로컬 GPU·HPC가 선택적으로 동원된다. ④ **데이터·장비 계층**은 기관의 지식자산과 실험 환경으로, 제어면을 통해서만 읽고 쓰인다.

계층	역할	주요 구성
① 사용자·업무	자연어 요청 접수, 결과·승인요청 반환	웹·문서·이메일·포털·랩노트
② Orchestration	민감도 판단·라우팅·도구 호출·승인·감사	정책 라우터, Skill, Memory, MCP, Agent workflow
③ 모델·계산	작업별 실행기(제어면이 호출)	글로벌 SOTA, EXAONE·HyperCLOVA X·Solar, 로컬 모델, HPC
④ 데이터·장비	기관 지식·실험환경(제어면 경유 접근)	내부 DB, Git, 장비망, 자율실험실, Run Ledger

5.2 각 연결의 의미

그림 1의 화살표는 다음을 뜻한다. 이 연결 규칙이 새결AI의 핵심 설계다.

연결	의미
① 사용자 ↔ ② 제어면	사용자가 자연어로 요청하고, 제어면은 결과 또는 ``승인 필요'' 신호를 돌려준다.
② 제어면 ↔ 사람 승인	외부 전송·장비 구동·특허성 산출처럼 위험이 큰 작업은 사람의 승인을 받은 뒤 진행한다(Human Gate).
② 제어면 → ③ 모델·계산	작업의 민감도·난도·비용을 따져 글로벌(공개·고난도)·국내 독자(소버린)·로컬 GPU(민감)·HPC(계산·장비) 중 하나로 라우팅한다.
② 제어면 ↔ ④ 데이터·장비	내부 데이터는 제어면이 직접 읽되, 외부로 나가는 경로에는 마스킹·추상화를 적용하고 모든 접근을 Run Ledger에 기록한다.

핵심은 **③ 모델과 ④ 데이터가 직접 연결되지 않는다**는 점이다. 예컨대 내부 실패 실험 로그(④)를 글로벌 SOTA AI(③)로 보내려면 반드시 제어면(②)을 거쳐 민감 항목이 가려지므로, “내부 문서가 외부 AI로 직행”하는 경로는 구조적으로 차단된다.

5.3 핵심 기능

제어면(②)은 다음 여섯 기능으로 구현된다.

- **민감도 기반 라우팅.** 입력·파일·도구 호출의 민감도를 분류해 외부·국내·로컬·TEE·사람 승인 중 실행 경로를 자동 선택한다 [1].
- **Mask & Hydrate.** 외부 AI에는 민감 항목을 placeholder·추상 질의로만 보내고(Mask), 외부 응답을 내부에서 원문과 결합해 복원한다(Hydrate). 핵심 IP를 내보내지 않으면서 외부 지식을 활용하는 장치다.
- **기관 Memory.** 기관의 결정·맥락·선례를 누적해, 사람이 바뀌어도 “업무 맥락”이 보존된다.
- **Skill 패키지.** 반복 업무·절차를 재사용 가능한 Skill로 묶어 day-one 인수인계를 지원한다.
- **MCP / 도구 연동.** 표준 프로토콜로 내부 DB·Git·장비망·자율실험실을 안전하게 연결한다.
- **감사 원장(Run Ledger).** 어떤 데이터를 어떤 모델·도구로 어떻게 처리했는지 전 과정을 기록해 재현성·감사·IP 증거를 남긴다 [20].

6. 대한민국 독자 파운데이션 모델 활용 필요성

새결AI는 글로벌 SOTA를 배제하지 않는다. 공개·저위험·고난도 작업에는 글로벌 AI가 유용하다. 그러나 국가 R&D·공공기관·대학·병원의 핵심 업무를 모두 해외 API에 의존하는 것은 데이터 주권 · 언어·제도 적합성 · 비용 안정성 · 기술 자립 측면에서 바람직하지 않다.

6.1 모델 활용 전략 --- 경쟁이 아니라 배치

새결AI는 모델을 경쟁 관계가 아니라 적재적소에 배치한다. 1M 컨텍스트와 다단계 추론 안정성이 필요한 **오케스트레이션**은 현 시점 글로벌 모델이 우위이나, 충분한 장문 컨텍스트를 지원하는 국내 독자 모델(예: EXAONE 계열)은 **Worker** 역할을 수행할 수 있다. 온프레미스·국내 추론 환경은 데이터 주권을 준수한다.

모델 유형	장점 / 한계	권장 역할	데이터 주권
글로벌 초거대 AI	최고 추론·1M 컨텍스트 / 반출·비용·의존	공개 자료, 오케스트레이션 상한	외부
국내 독자 모델	주권·한국어 적합 / 일부 도구사용 확인 필요	기관 업무 Worker	국내
오픈소스 로컬	온프레미스·커스터마이징 / GPU·튜닝 필요	민감·반복 작업	온프레미스
도메인 특화	소재·바이오 성능 / 범용 추론 한계	전문 분석 Worker	가변

7. Scientific Discovery는 반복 탐색 문제이다

과학 발견은 한 번의 정답 생성이 아니라 반복이다. 이 반복은 트리 탐색과 동형이며, 새결AI의 **Search & Optimization Control Plane**이 이를 운영한다.

7.1 결합 획득함수: 기존 연구와의 차별점

기존 자율실험 폐루프와 LLM 기반 추론 탐색은 대체로 성능 향상 가능성(보상)만으로 다음 단계를 선택한다 [3,29]. 새결AI는 여기에 비용·누출·IP를 결합한 획득함수로 다음 후보 x와 실행 경로 route를 동시에 선택한다.

$$a(x, \text{route}) = \underbrace{\mathbb{E}[\Delta\text{Perf}(x)]}_{\text{정보이득}} - \lambda_c \,\text{Cost}(x) - \lambda_\ell \,\text{Leak}(x, \text{route}) - \lambda_{ip} \,\text{IPRisk}(x).$$

여기서 Leak은 실행 경로에 따른 정보 누출 위험, IPRisk는 후보의 특허 가치 노출, λ는 기관 정책 가중치다. 이는 입력 민감도로 실행 경로를 분기하는 PRISM [1]을 "탐색 단계 선택" 수준으로 확장한 것이다. 다만 LLM 단독 탐색이 통계적 기법을 항상 능가하지는 않으므로 [38,39], 본 시스템은 GP / 베이지안 surrogate와 LLM-guided 탐색의 **하이브리드**를 전제하며, 불확실성을 반영한 MCTS [32] 위에 위 가치함수를 결합한다.

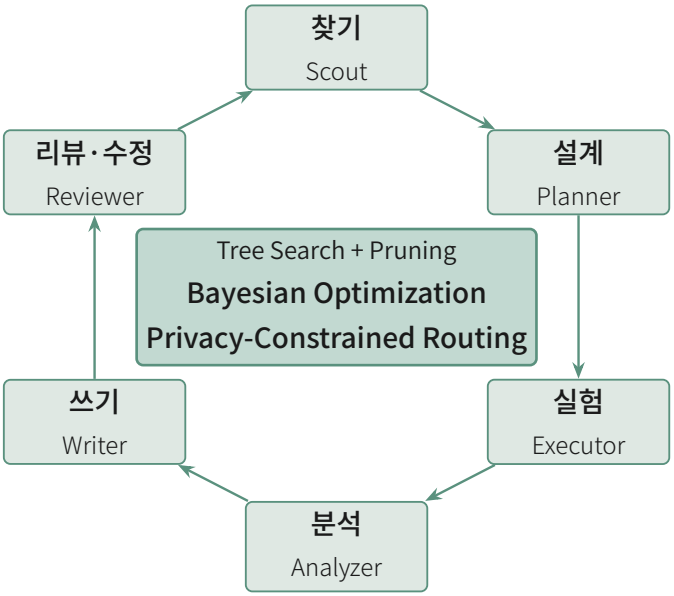

Figure 2. Scientific Discovery MissionLoop. 찾기–설계–실험–분석–쓰기–리뷰–수정의 폐루프에서, 후보 가설·실험 경로가 branch가 되고 실험·검토 결과가 branch의 점수가 된다. 유망하지 않은 branch는 pruning하고 유망한 branch에 계산·실험 예산을 배정한다.

7.2 탐색은 학습된 평가기로 가지치기된다 --- 강화학습·MCTS의 작동 원리

거대한 탐색공간을 학습된 평가기로 안내하고 가지치기하는 패러다임은 강화학습에서 이미 검증되었다. 강화학습은 네 가지가 맞물려 도는 폐루프다. 에이전트는 현재 상황에서 어떤 행동을 할지 알려 주는 **정책**을 따르고, 각 상황이 앞으로 얼마나 좋은 결과로 이어질지를 가늠하는 **가치 평가기**를 가지며, 실제로 행동하며 얻은 **경험**(상황·행동·보상의 기록)을 쌓고, 그 경험으로 정책과 가치 평가기를 다시 더 똑똑하게 만든다. 핵심은 더 나은 평가가 더 나은 행동을 낳고, 더 나은 행동이 더 풍부한 경험을 만들어, 시스템이 스스로 점점 강해진다는 점이다.

Atari 게임 --- 경험만으로 스스로 더 잘하기. DQN [60]은 게임 화면(픽셀)만 보고 점수라는 보상만으로 학습한다. 에이전트는 같은 게임을 수없이 반복하며 "이 상황에서 이렇게 움직이면 결국 점수가 얼마나 오르는가"를 조금씩 더 정확히 가늠하게 되고, 그 가늠이 좋아질수록 더 나은 수를 두며, 더 나은 플레이는 다시 더 좋은 경험을 만든다. 사람이 전략을 가르쳐 주지 않았는데도 이 되먹임만으로 49개 고전 게임에서 숙련된 사람 수준에 도달했다. 다만 DQN은 여러 수 앞을 내다보는 탐색 없이 그 순간의 최선을 고르는 반응형이다.

바둑 --- 앞을 내다보는 탐색과 자기대국. 바둑은 둘 수 있는 경우의 수가 천문학적이어서 모든 수를 다 따져 볼 수 없다. AlphaGo [45]는 두 가지 "직관"을 학습해 따져 볼 양을 줄였다. 하나는 "지금 둘 만한 수"로 후보를 좁혀 주어 쓸데없이 넓게 보지 않게 하는 **정책 직관**이고, 다른 하나는 끝까지 두어 보지 않고도 "이 국면이 유리한가"를 한눈에 가늠해 쓸데없이 깊게 보지 않게 하는 **가치 직관**이다. 몬테카를로 트리 탐색은 이 두 직관을 길잡이 삼아, 유망한 길은 더 깊이 들여다보고 가망 없는 길은 일찍 접으며 몇 수 앞을 내다본다. 결정적으로, 이렇게 "내다보며 고른 수"는 원래의 정책 직관 하나보다 더 강하므로, AlphaGo는 그 결과를 자신의 새 교과서로 삼아 두 직관을 다시 학습한다. AlphaZero [46]는 사람의 기보 없이 무작위 상태에서 출발해, 자기 자신과 두는 대국 $\rightarrow$ 더 똑똑해진 두 직관 $\rightarrow$ 더 강한 탐색 $\rightarrow$ 더 좋은 대국이라는 순환만으로 초인 수준에 이르렀다. "스스로 더 잘 두게 되는" 비결이 바로 이 순환이다.

AlphaTensor --- 같은 방법을 과학적 발견으로. 연구진은 "더 빠른 행렬곱셈 알고리즘 찾기"를 한 사람이 푸는 퍼즐 게임처럼 바꾸었다 — 한 수 한 수가 알고리즘의 한 조각이고, 더 적은 연산으로 정답에 이르면 이기는 게임이다. 바둑과 똑같은 탐색 기계로, 사람이 50년간 넘지 못한 4×4 행렬곱 기록(Strassen)을 경신했다 [47]. "다음 한 수 = 다음 실험"으로 바꾸면 자율실험과 구조가 같다.

AlphaFold2 --- 탐색 대신 뛰어난 예측기. 대조적으로 단백질 구조 예측의 AlphaFold2 [48]는 앞을 내다보는 탐색을 쓰지 않는다. 방대한 진화 정보를 학습한 예측기가 구조를 단번에 내놓는다. 즉 평가·예측기가 충분히 정확하면 탐색 없이도 강력하며, 거꾸로 탐색의 성패도 결국 평가기의 품질이 좌우한다 [35].

새결AI에의 함의. 새결AI의 후보 트리 생성기는 바둑의 정책 직관처럼 유망한 후보로 폭을 좁히고, 도메인 예측기(surrogate)는 가치 직관처럼 실험 없이 후보의 전망을 가늠해 깊이를 줄이며, 자율실험 결과는 경험이 되어 그 예측기를 갱신한다. 본 제안의 차별점은 이 평가기에 **프라이버시·IP 위험**을 함께 담아(7.1절), "성능만"이 아니라 "성능과 보호"를 동시에 최적화하는 탐색을 만든다는 데 있다.

7.3 PI의 최적화 전문성 --- Sphere Decoding = 잡음 하 탐색 트리 가지치기

"다음 후보 선택"과 "다음 추론 경로 선택"은 본질적으로 동일한 트리 탐색 + 획득 / 가치 함수 문제다 [41]. 제안 연구진은 이 문제의 수학적 핵심인 잡음 하 고차원 조합 탐색공간의 가지치기에 장기 전문성을 보유한다.

MIMO 최대우도(ML) 검파는 정수 최소자승(integer least-squares) 문제

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{C}^N} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$$

로, 후보 수가 $|\mathcal{C}|^N$로 지수 증가한다. **Sphere decoding**은 $\mathbf{H} = \mathbf{QR}$ 분해 후 상삼각 **R**을 이용해 부분 거리(partial Euclidean distance)를 깊이순으로 누적하는 깊이우선 분기한정(branch-and-bound) 탐색이다. 어떤 부분경로의 누적 거리가 반경 d를 넘으면 그 서브트리 전체를 즉시 가지치기한다. 이는 정확히 "유망하지 않은 가지를 경계(bound)로 잘라내는" 탐색트리 가지치기이며, 잡음(**y**에 섞인 채널 잡음) 하에서 이산 후보를 고르는 의사결정이다. 따라서 잡음 섞인 실험 피드백 하에서 다음 실험 / 후보를 고르는 문제와 구조적으로 동형이며, 부분거리 경계는 새결AI 획득함수의 가치 하한 역할에 직접 대응한다.

- **Circular Sphere Decoding (IEEE TVT 2017).** 임의 2D 성상도에 대한 저복잡도 복소 sphere decoding을 제안하고, 특정 탐색트리의 가지치기 용량(pruning capacity)을 추정하는 도구와 이를 활용한 **Predict-and-Change** 전략으로 추가 복잡도를 절감했다 [49] — 어떤 가지를 얼마나 잘라낼 수 있는지를 예측해 탐색을 설계하는 직접적 방법론이다.
- **Tree-Pruning Detection & LDPCC Decoding (IEEE JSAC 2005), MIMO over Fading (IEEE TCOM 2008).** sphere list detection과 임계값 기반 tree-search를 결합한 반복 가지치기 검파·복호 경험 [50,51]은, 잡음 하에서 살아남은 경로 집합을 관리하는 문제로서 Candidate Tree Builder의 가지치기 모듈과 noisy acquisition 설계로 곧장 이어진다.

요컨대 AlphaGo류가 "가치망으로 가지치기"한다면, sphere decoding은 "거리 경계로 가지치기"한다 — 둘은 같은 원리이며, PI의 검파 / 복호 전문성이 새결AI 탐색 엔진의 수학적 토대가 된다.

7.4 최신 Bayesian / Tree-Search 최적화 계보

7.2–7.3이 "탐색은 학습된 평가기로 가지치기된다"는 원리였다면, 본 절은 그 원리가 LLM 시대에 어떻게 구체화되었는지를 정리한다. 흐름은 세 갈래가 한 점으로 수렴하는 것으로 읽으면 이해가 쉽다.

(1) 추론을 트리로. 초기 LLM은 한 줄로 "생각"했지만, ToT·RAP [29,30]는 부분해를 노드로 하는 **트리**를 펼쳐 여러 추론 경로를 비교·역추적한다. 이는 7.2의 게임 트리와 같은 구조다. (2) 트리를 가치로 가지치기. 트리를 무작정 넓히면 비용이 폭증하므로, ReST-MCTS*·PlanU·LATS [31,32,34]는 각 단계의 **가치(process reward)**를 학습·추정하고 불확실성을 반영한 MCTS로 유망한 가지에만 예산을 쓴다 — AlphaGo의 정책·가치망이 LLM 추론에 이식된 셈이다. (3) 적은 실험으로 후보 고르기. 동시에, 실험 / 평가 비용이 큰 과학 문제에서는 **베이지안 최적화(BO)**가 surrogate

와 획득함수로 다음 후보를 고른다. LLAMBO·GFlowNet [36,40]는 여기에 LLM의 사전지식을 결합해 cold-start를 줄이고 다양한 후보를 배치 샘플링한다.

기법(인용)	핵심	새결AI 적용
ToT [29], RAP [30]	추론을 트리 탐색 / world-model planning으로 정식화	추론 경로 탐색의 기준점
ReST-MCTS* [31], PlanU [32]	단계 가치 추정·불확실성 반영 MCTS	노이즈 하 후보·경로 선택
LATS [34], MCTS-AHD [33]	추론·행동·휴리스틱 설계 통합 탐색	Agent 계획·실험전략 자동설계
LLAMBO [36], GFlowNet [40]	LLM-guided BO·다목적 후보 배치 샘플링	후보 ranking / batch 선택
Sober look [39], Are we there [38]	LLM 단독 BO의 한계·GP 백본 필요	하이브리드 전제의 근거
SyntheMol [42], MC Thought Search [41]	합성가능 분자·촉매를 MCTS로 탐색	후보물질 탐색의 동형 실증

중요한 단서는 세 번째 흐름에 대한 비판이다. Sober look·"Are we there" [38,39]는 LLM 단독 BO가 잘 조율된 통계적 surrogate(GP)를 항상 능가하지는 못함을 보였다. 그래서 새결AI는 "LLM이 후보를 제안하고, GP / MCTS가 검증·선택"하는 **하이브리드**를 전제한다 — LLM은 도메인 지식으로 탐색 공간을 좁히고, 통계적 엔진은 적은 실험으로 신뢰성 있게 다음 점을 고른다. SyntheMol·MC Thought Search [41,42]는 이 결합이 분자·촉매 탐색에서 실제로 작동함을 보인 동형 사례다.

시사점. 세 흐름의 공통 결론은 "좋은 탐색은 좋은 평가기(가치함수 / surrogate)에서 나온다"는 것이다. 따라서 새결AI의 경쟁력은 더 큰 모델이 아니라 도메인 평가기 설계에 있으며, 본 제안은 여기에 한 축을 더한다 — 평가기에 **프라이버시·IP 위험**을 내장(7.1절 획득함수)하여, "성능만"이 아니라 "성능과 보호"를 함께 최적화하는 탐색을 만든다. 이는 위 계보 어디에도 없는 부분이며 새결AI의 핵심 신규성이다.

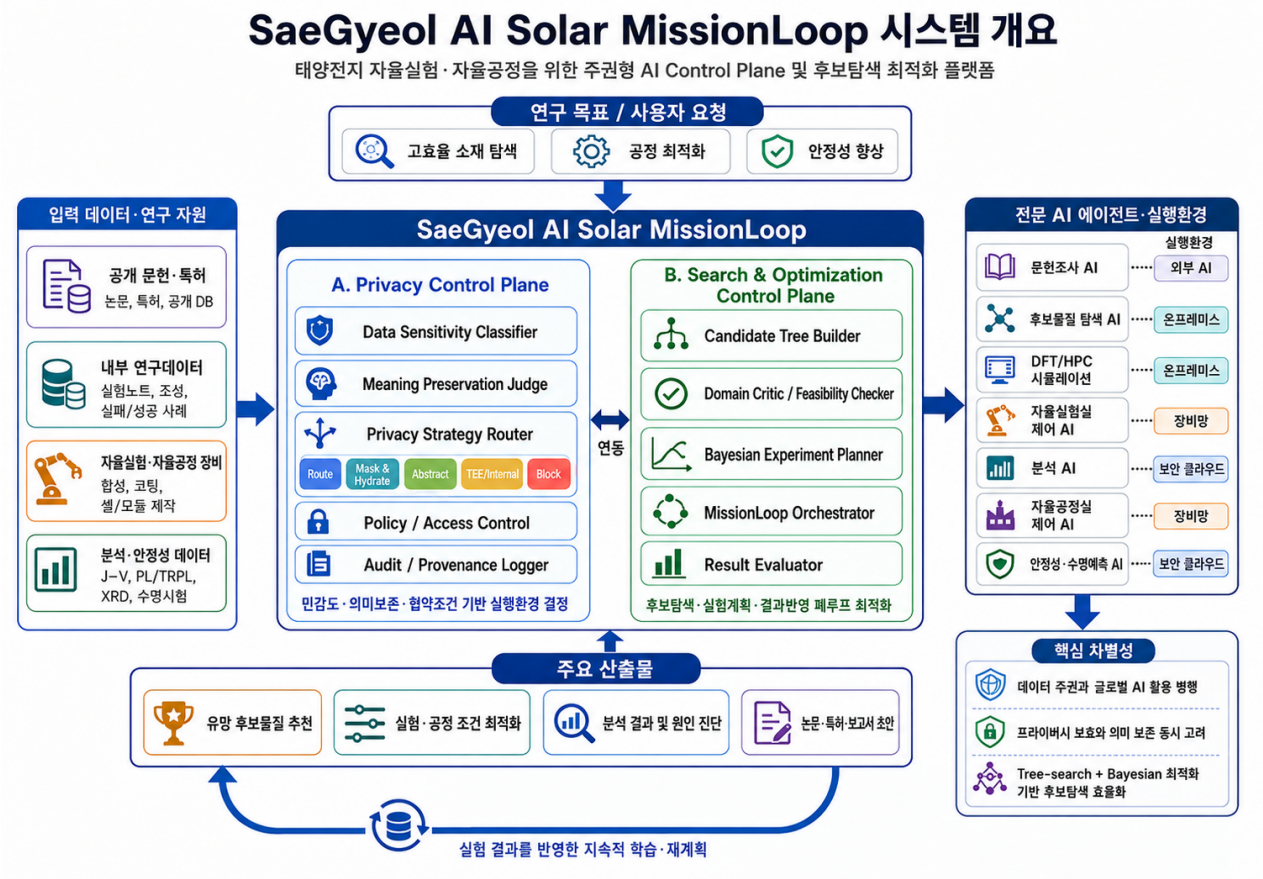

Figure 3. 새결 Private AI Solar MissionLoop 시스템 개요(그림 3). 좌측 Privacy Control Plane(민감도 분류·의미보존 마스킹·라우팅·감사)과 우측 Search & Optimization Control Plane(Candidate Tree Builder·Bayesian Experiment Planner·결과 평가)이 연동되어, 공개 작업은 외부 AI로 핵심 IP는 내부·TEE로 라우팅하면서 후보탐색을 폐루프로 최적화한다.

8. 1차 적용 사례: 새결 Private AI Solar MissionLoop

페로브스카이트/탠덤 태양전지는 소재 탐색 · 박막 공정 · 셀 제작 · 모듈화 · 장기 안정성 평가가 모두 반복 루프로 구성되어 AI 과학자 실증에 적합하다. 실제로 Gao 연구진은 능동학습·양자모델링으로 소재를 탐색하고 베이지안 최적화·기호회귀로 제조 공정을 정련하는 자율 폐루프를 구축해, 패시베이션 분자 5ANI로 0.05 cm^2 셀에서 PCE 27.22%(인증 MPPT 27.18%), 21.4 cm^2 미니모듈에서 23.49%를 달성했고, ISOS-L-1I 1,200시간 후 초기효율의 98.7%를 유지했으며 수동 제작 대비 약 5배의 재현성을 보였다 [3]. 이는 태양전지 R&D가 “후보 생성–실험–측정–다음 후보 선택”의 반복 최적화 문제로 정식화됨을 입증한다.

8.1 태양전지 R&D 병목과 Solar MissionLoop 구조

그림 3은 5절의 일반 구조를 태양전지에 특화한 것이다. 왼쪽 절반(**Privacy Control Plane**)은 입력 데이터를 민감도로 분류하고 의미를 보존하며 마스킹한 뒤 실행 경로를 정하고 모든 처리를 감사 기록한다. 오른쪽 절반(**Search & Optimization Control Plane**)은 후보를 트리로 생성하고(Candidate Tree Builder), 정보이득과 실험비용·IP 위험을 함께 고려해 다음 실험을 계획하며(Bayesian Experiment Planner), 결과를 평가해 다음 라운드로 되먹인다. 두 제어면은 가운데에서 연동되어, 공개·저위험 작업은 외부 AI로, 핵심 IP는 내부·TEE로 보내면서 후보 탐색을 하나의 폐루프로 최적화한다. 아래 표는 각 구성요소의 역할이다.

계면·결함 제어가 성능을 좌우하는 고차원 설계 문제이고 [52,53], 모듈화·장기 안정성은 별도의 공정·데이터 문제이며 [54], 실패 조건·수율·공정 window는 핵심 IP다. Solar MissionLoop는 이를 공통 객체모델·민감도 라우터·후보 트리·프라이버시-제약 실험계획·자율실험 연동·감사 원장으로 구성한다.

구성요소	역할
Solar Mission Object Model	샘플·조성·후보 분자·공정조건·측정·장비로그·안정성 곡선을 공통 스키마로 정의
Privacy Strategy Router	공개 / 민감 / 고위험 데이터를 Route·Mask·Abstract·Internal / TEE·Block 중 선택
Candidate Tree Builder	후보 분자·작용기·공정조건을 트리로 생성하고 가지치기 [29,42]
Privacy-Constrained Experiment Planner	정보이득·실험비용·IP 위험을 함께 고려해 다음 batch 선택
Run Ledger & Evidence Logger	데이터·모델·프롬프트·라우팅 결정·후보 탈락 이유 기록 [20]

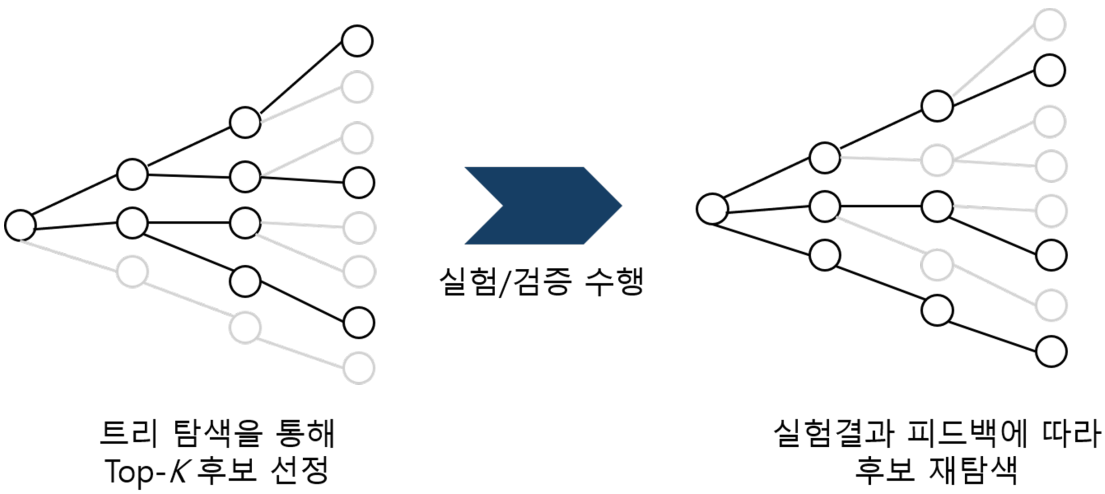

Figure 4. 실험 결과에 따른 Tree-search 피드백(그림 4). 트리 탐색으로 Top-K 후보를 선정 → 실험 / 검증 수행 → 결과 피드백에 따라 유망 branch를 갱신하고 비유망 branch를 가지치기하여 후보를 재탐색한다.

8.2 프라이버시-제약 실험설계

일반 Bayesian optimization [28]은 “어떤 실험이 성능을 가장 높일 가능성이 큰가?”만 묻는다. 새결AI는 “그 실험을 계획·분석하기 위해 어떤 데이터를 어떤 AI에 보내도 되는가?”를 함께 묻는다. 즉 7.1절 획득함수의 항(예상 성능 향상, 불확실성 감소, 실험 비용, 정보 누출 위험, 정책 제약)을 동시에 평가해 다음 실험과 실행 경로를 선택한다.

8.3 Privacy 심화: PRISM을 "구조화 과학 산출물"로 확장

PRISM [1]은 사용자가 입력한 자연어 문장의 민감도를 보고 클라우드·엣지·협업 중 경로를 정하는 라우팅 기법이다. 그러나 연구 현장에서 정작 보호해야 할 것은 문장이 아니라 **구조화된 과학 데이터**이다. 새결AI는 보호 대상을 분자 구조·조성·공정 window·실패 로그·안정성 곡선 같은 "구조화 과학 산출물"로 넓힌다. 이것이 8.3의 핵심 확장이다.

이때 다루어야 할 위협은 네 가지다. (i) 데이터를 외부 AI에 업로드하면서 영업비밀이 새는 경우, (ii) 모델 학습 / 파라미터(gradient)를 통해 원본 데이터가 역추론되는 경우, (iii) 여러 기관이 공동연구할 때 데이터 / 성과의 소유권이 침해되는 경우, (iv) Agent가 만든 산출물의 발명자·권리 귀속이 불명확해지는 경우다 [14,25].

대응은 세 층위로 구성한다. 기술 층위에서는 연합학습·secure aggregation·split learning·차분 프라이버시(DP)·신뢰실행환경(TEE)을 데이터 성격에 맞게 조합한다 [12,13,14,15,16]. 기록 층위에서는 W3C PROV 표준으로 "무엇을·누가·어떤 모델로 처리했는가"의 출처(provenance)를 남긴다 [20]. 거버넌스 층위에서는 NIST AI RMF / 생성형 AI 프로파일 / 프라이버시 프레임워크 / Zero Trust로 운영 규범을, USPTO 발명자성 지침과 EU AI Act로 IP·규제 정합을 확보한다 [21,22,23,24,25,26,27]. 아래 박스는 이 원리가 실제 한 번의 분석에서 어떻게 작동하는지 보여 준다.

> **4단계 Mask → Descriptor → External → Hydrate 예시.** (가정) 신규 passivator 후보가 초기 PCE는 높았으나 고온 light-soaking에서 빠른 burn-in 발생. **(1) Mask** --- 연구자명·샘플 ID·후보명·조성·공정조건을 placeholder로 치환. **(2) Descriptor** --- 정확한 구조·수치는 제거하되 "pyridine / cyano 계열 작용기", "고온 광스트레스 하 early burn-in" 등 분석에 필요한 의미만 보존해 외부 질의문 구성. **(3) External** --- 외부 AI가 민감정보 없이 공개 지식 기반 열화 메커니즘(deep-trap 패시베이션 미흡·이온 이동·계면 반응·봉지 불량)과 대조 실험(PL/TRPL, TOF-SIMS, O_2/습도 분리 light soaking)을 제시. **(4) Hydrate** --- 외부 응답을 내부 원본(실제 후보명·조건)과 결합해 실행 가능한 다음 실험 순서로 구체화. 핵심 IP는 외부로 나가지 않으면서 외부 지식은 활용된다.

8.4 선행 연구와의 차별성 및 연구질문

새결AI의 신규성은 개별 요소가 아니라 자율 폐루프 + 프라이버시 라우팅 + 다기관 거버넌스의 교집합에 있다.

선행 접근	한계	새결AI 확장
자율실험 폐루프 [3,55]	단일 연구실·단일 데이터 흐름, 데이터 주권 미반영	다기관 민감도 라우팅 + 감사 원장 결합
프라이버시 라우팅 [1]	자연어 프롬프트의 의미 민감도 중심	구조화 과학 산출물의 민감도·역추론 위험 반영
모델·비용 라우팅 [17,18]	비용·성능 위주, 프라이버시·폐루프 결합 부족	외부·내부·HPC·장비망·TEE를 워크플로 단위로 선택
새결AI(본 제안)	---	위 세 축을 결합한 프라이버시-제약 폐루프 실험설계

연구질문(RQ)

- **RQ1 (민감도)** 구조화 과학 산출물의 민감도를 자동으로 프로파일링하고, 외부 노출 시 역추론 위험을 정량화할 수 있는가?

- **RQ2 (보장)** 의미를 보존하는 마스킹의 "유용성–누출" 트레이드오프를 측정하고 보장할 수 있는가?
- **RQ3 (결합최적화)** 정보이득과 누출을 동시에 고려한 획득함수 / 라우팅이 단일 목적 방식보다 우월한가?
- **RQ4 (거버넌스)** 정책·승인·감사 원장으로 처리의 감사가능성과 재현성을 보장할 수 있는가?

검증가설(H)

- **H1** 구조화 민감도 평가가 단순 개체명인식(NER)보다 정보 누출을 낮춘다.
- **H2** "유용성–누출" 공동최적 라우팅이 전부-내부(품질 손해)와 단순-마스킹(누출 증가)을 모두 능가한다.
- **H3** 프라이버시-제약 Bayesian optimization이 동일한 IP 예산에서 더 적은 실험으로 목표 성능에 도달한다.

각 가설은 12절의 정량 지표(민감 span F1, 누출률, 목표 도달 실험 횟수)로 검증한다.

9. K-문샷 사업과의 정합성

9.1 국가 과학AI 통합플랫폼에서 새결AI의 위치

K-문샷은 2030년까지 AI 기반 연구생산성 2배, 2035년까지 8대 분야 12대 국가 미션 해결을 목표로 하며, 국가과학AI연구센터(NAIS)를 중심으로 연구데이터·GPU·AI 모델·자율실험실·AI 에이전트의 5대 자원을 통합한다 [56]. 새결AI는 이 중 **AI 에이전트**와 **Private Orchestration Layer**를 담당한다.

9.2 12대 미션 공통 적용 --- GIST 직접 정합 8개 미션

태양전지는 1차 적용 사례이며, 동일 구조가 GIST 연구소 기반의 다른 미션으로 확장된다(참고: GIST 주도 K-문샷 공공연구 사업계획 [58]).

미션	GIST 연구기반	새결AI 적용(요지)
핵융합	고등광기술연구소	실험데이터 분석, 고출력 레이저 조건 정리, 중성자·입자신호 해석
AI 과학자	인공지능·중앙기기연구소	가설 생성, 실험계획, 장비데이터 해석, 보고서 작성
AI 반도체	첨단AI반도체팹센터	공정조건 추천, 패키징·신뢰성 데이터 관리, 검증 문서화
첨단바이오·신약	생명의과학융합·실험동물자원센터	표적 탐색 보조, 오믹스·전임상 정리, 실험계획서 작성
BCI	생명의과학융합연구소	뇌신호 분석, 피험자·조건 관리, 해독모델 학습데이터 구축
태양전지	차세대에너지연구소	소재·구조 탐색, 성능·열화 분석 (1차 실증)
피지컬 AI	인공지능·팹·중앙기기	센서·장비 연계, 물리실험 예측, 온디바이스 로그 분석
우주 데이터센터	팹·차세대에너지연구소	우주용 반도체·태양광·내환경 소재 데이터 정리

최적화(7절)와 Privacy Routing(8절)은 어느 미션에서나 필요한 공통 Core이므로, 새결AI는 "12대 미션 공통 후방 AI 도구층"으로 일반화된다.

9.3 K-문샷 고유 목적 달성 지원

새결AI는 (i) 반복 탐색·실험계획을 Agentic MissionLoop로 가속하여 연구생산성 향상에, (ii) 국내 독자 모델·로컬·국가 GPU를 하나의 Control Plane으로 연결하여 데이터 주권 확보에, (iii) Run Ledger로 재현성에, (iv) 태양전지 검증 구조의 타 미션 확산으로 분야 확장성에 기여한다.

10. 단계별 추진계획

본 사업은 GIST 주도 “데이터 주권형 범용 Agentic AI 플랫폼 개발”로서 **2026--2030(5년)** 기간을 상정하며 [58], 마일스톤은 활동이 아니라 측정 가능한 성과로 기술한다. (사업비는 추진단 협의 대상)

단계	성과 목표(outcome)	산출물
1. 설계 / PoC	민감 span F1 $\geq$ 0.90, 라우팅 정확도 $\geq$ 0.90, 정책 위반 0건	v1 플랫폼, 벤치마크 질의셋, 후보탐색 데모
2. 태양전지 페루프	동일 IP예산 하 목표 성능 도달 실험 횟수·DFT 호출 절감	Solar MissionLoop, 후보 우선순위표, 실험 원장
3. 공정 최적화	batch 재현성 향상, 장비 제약 위반 0건	Recipe card, 자율공정 최적화 demo
4. 다기관 실증	출연연·대학·기업 3--5곳 적용, 비식별 패키지 검증	권한관리·감사 체계, 비식별 데이터 패키지
5. 국가·산업 확산	12대 미션 후방 AI 도구층·NAIS 연계	MissionLoop template, 도메인 skill pack

PoC 규모(3--6개월). 사용자 50명 내외, 온프레미스 GPU(80 GB급 다수 또는 동등 기관 GPU), 제한적 외부 API 사용, 미션 PD 1인 연계. 데이터는 공개 논문·비식별 실험노트·샘플 스키마·일부 내부 문서.

10.1 주요 위험과 완화

위험	완화 방안
국내 독자 모델 단독 오케스트레이션의 안정성 미검증	글로벌 모델 오케스트레이션 + 국내 모델 Worker 하이브리드로 시작, 단계적 이관·검증
민감도 오분류로 인한 정보 누출	red-team reconstruction 평가, human approval gate, 정책 위반 0건을 출시 기준으로 설정
자율실험 장비 안전·오작동	장비 제약 사전 검증, human-in-the-loop 승인, 실행 원장 기록
데이터 반출·개인정보 규제	K-문샷 특별법 데이터 특례 활용, 비식별·마스킹, 감사 원장 기반 추적
외부 AI 의존(가격·약관·중단)	다중 라우팅·비용 cap·국내 대체 경로로 단일 공급자 종속 회피

10.2 추진 체계와 역할

역할	담당 범위
PI / 총괄	연구 방향·미션 정합·대외 협력·의사결정
플랫폼 팀	제어면·민감도 라우팅·Mask&Hydrate·Memory·감사 원장 개발
탐색 / 최적화 팀	Candidate Tree Builder·surrogate·프라이버시-제약 획득함수
도메인 팀(태양전지·EECS)	MissionLoop 구성·자율실험 연동·질의셋 구축
평가 / 레드팀	벤치마크·red-team reconstruction·지표 측정
거버넌스 / IP	정책·승인 흐름·provenance·발명자성·규제 정합

10.3 단계 의존성

단계는 독립적이지 않고 순차 의존한다. **1(설계 / PoC)** — 제어면·평가 하니스를 먼저 확보해야 후속 단계가 그 위에서 동작한다. **2·3(태양전지 페루프·공정 최적화)** — 1의 제어면·탐색 엔진에 의존하며, 도메인 데이터·장비 연동이 선행 조건이다. **4(다기관 실증)** — 2·3에서 검증된 민감도·라우팅·비식별 패키지가 전제다. **5(국가·산업 확산)** — 4의 권한·감사 체계와 도메인 Skill pack을 재사용해 타 미션으로 복제한다. 핵심 임계경로(critical path)는 제어면 → 태양전지 페루프 → 다기관 권한 / 감사이며, 평가 하니스는 전 단계와 병렬로 진행된다.

11. 예산 구조와 PoC 산출물 계획

본 사업의 투자는 성격이 다른 세 묶음으로 구분된다. 새결AI는 단독으로도 가치가 있으나, GIST 8개 미션 분야와 연결될수록 공통 후방 AI 도구층으로서 효용이 비선형적으로 커진다 — 같은 제어면·탐색 엔진을 여러 미션이 공유하기 때문이다. 따라서 예산도 "코어 개발"과 "미션 연계 확장"을 분리해 설계한다. (구체적 사업비·단가는 추진단 협의 대상)

11.1 투자 구조(정성)

묶음	내용	성격
A. 인프라 구축	소형 데이터센터 및 GPU 인프라(온프레미스 추론·자율실험 연계 기반)	초기 자본투자(1회성)
B. 새결AI 코어 R&D	제어면·Privacy Routing·탐색 / 최적화 엔진·감사 원장 개발(다년)	지속 R&D
C. 8개 미션 연계·실증	핵융합·AI과학자·반도체·바이오·BCI·태양전지·피지컬AI·우주 분야와 연결·실증	확장(연계 시 효용 급증)

핵심 원칙은 A·B로 만든 하나의 코어를 C에서 재사용한다는 점이다. 미션이 늘어도 제어면·탐색 엔진은 공유되고 도메인 Skill / 데이터 스키마만 추가되므로, 미션당 한계비용이 낮아지고 투자 대비 효과가 커진다.

11.2 2026년 PoC 계획과 연말 산출물

첫 해는 “데이터 주권형 자율연구 워크플로가 실제로 동작함”을 보이는 데 집중한다. Local DGX 서버 1세트를 구축하고 새결AI를 설치한 뒤, Science AI 자율실험을 서로 다른 두 분야에서 시연한다.

항목	내용
인프라	Local DGX 서버 1세트 구축, 새결AI 시스템 설치·연동
시연 1 — EECS 분야	전자·통신 연구의 Agentic 워크플로(문헌 → 가설 → 코드 → 분석 페루프) 1식
시연 2 — 태양전지 분야	Solar MissionLoop 기반 후보탐색·실험계획 자율 워크플로 1식
연말 산출물(Deliverable)	두 분야 Agentic AI workflow 시연 결과를 바탕으로 본 사업의 Deliverable·확장 범위·다음 단계 지표를 확정

즉 2026년 PoC의 성공 기준은 “보고서”가 아니라 두 개의 서로 다른 자율연구 워크플로가 새결AI 위에서 실제로 도는 것이며, 그 결과로 후속 단계의 deliverable을 결정한다.

12. 평가체계 및 성공 기준

인상평가가 아니라 정량 하니스로 검증한다. 태양전지 질의셋을 공개·준공개·내부·핵심 IP·공동연구·민감 공정의 6개 유형으로 구성하고 민감 span·정답 라우팅 라벨을 부여하며, 4개 베이스라인(All-Internal·All-External·Naive-Masking·PRISM-style)과 비교한다.

평가축	지표	목표
작업 품질	전문가 평가·LLM-judge, 다음 실험 추천 적중	All-External 대비 90% 이상 유지
프라이버시	red-team reconstruction 누출률, 정책 위반 건수	Naive-Masking 대비 누출 50%↓, 위반 0건
민감도 판단	민감 span F1, 라우팅 정확도	F1 $\geq$ 0.90, 라우팅 $\geq$ 0.90
효율	완료시간·비용·에너지	All-Internal 대비 지연·비용 동등 이하
폐루프 효율	목표 성능 도달 실험 횟수, DFT 호출 수	동일 IP예산 하 절감

12.1 예시 결과(합성 데이터 기반 잠정치)

아래는 설계 검증을 위한 합성·익명화 질의셋 기반의 **예시(illustrative) 잠정 수치**이며, 실제 파일럿 데이터로 대체될 예정이다. 절대값이 아니라 베이스라인 간 상대 경향을 보기 위한 것이다.

지표	All-Ext.	All-Int.	Naive-Mask	PRISM	새결AI
민감 span F1 ↑	---	---	0.74	0.86	**0.93**
누출률(red-team) ↓	1.00	0.00	0.34	0.19	**0.12**
작업 품질(Ext. = 1.0) ↑	1.00	0.78	0.82	0.88	**0.94**
목표도달 실험 횟수 ↓	---	---	---	38	**23**
정책 위반 건수 ↓	다수	0	중간	소수	**0**

해석: 새결AI는 외부 전송 수준의 작업 품질(0.94)을 유지하면서, 누출률을 단순 마스킹 대비 약 65% 낮추고(0.34에서 0.12로), 동일 IP 예산에서 목표 성능 도달에 필요한 실험 횟수를 단순 BO 대비 약 40% 줄이는 것을 목표로 한다. 위 수치는 합성 데이터에 기반한 예시이며, 본 사업 1단계에서 실제 측정으로 확정한다.

12.2 추적성 매트릭스

연구질문(8.4절)–지표–마일스톤–산출물을 1:1로 연결해 검증 누락을 방지한다.

연구질문	가설	지표	마일스톤	산출물
RQ1 민감도	H1	민감 span F1, 역추론 위험	1	민감도 분류기, 질의 셋
RQ2 보장	H1	누출률, 정책 위반 건수	1--2	red-team 리포트
RQ3 결합최적화	H2, H3	작업 품질, 목표도달 실험 횟수	2--3	MissionLoop, 후보 우선순위표
RQ4 거버넌스	---	감사 완전성, 재현율	4	권한·감사 체계

13. 기대효과

관점	기대효과
연구기관	기관 지식자산 보존, 연구 생산성 향상, 데이터 주권 확보, 인수인계 비용 절감, 공동연구 신뢰성 향상
국가·K-문샷	독자 AI 생태계 강화, 연구 데이터 자산화, 12대 미션 공통 인프라 제공, 산업 확산, 소버린 AI 실증
방법론	“프라이버시-제약 폐루프 실험설계”라는 새로운 문제 정식화와 유용성--누출 평가 체계 제시

13.1 정량적 근거

기대효과는 다음 외부 지표로 뒷받침된다. (연구생산성) K-문샷은 2030년까지 AI 기반 연구생산성 2배를 공식 목표로 하며 [56,58], 새결AI의 MissionLoop는 후보탐색·실험계획 가속으로 이 목표에 직접 기여한다. (IP·데이터 보호의 가치) IBM 2025 조사에 따르면 데이터 유출의 글로벌 평균 피해는 약 $4.44M이고, 특히 미승인 외부 AI(shadow AI) 사용은 건당 약 $670K의 추가 비용을 유발하며 조사 기관의 63%가 AI 거버넌스 정책을 갖추지 못했다 [62] — 새결AI의 정책 기반 라우팅·감사 원장은 바로 이 “거버넌스 공백”을 겨냥한다. (시장) McKinsey는 소버린 AI를 2030년 약 $500–600B 규모로 추정하며 AI 지출의 30–40%가 주권 요건의 영향을 받을 것으로 본다 [61] — 새결AI는 이 흐름에서 국내 기관용 데이터 주권 인프라의 표준 후보가 될 수 있다.

14. 제안의 핵심 메시지

- **AI4Sci 시대에는 연구 자체가 거대한 반복 탐색 문제가 된다.**
- **외부 초거대 AI는 강력하지만, 기관의 지식자산과 업무 맥락을 외부에 맡길 수는 없다.**
- **새결AI는 모델 하나가 아니라, 민감도·비용·성능·정책을 함께 판단하는 Private AI Control Plane이다.**
- **국내 독자 파운데이션 모델과 로컬 GPU는 데이터 주권형 AI Scientist의 핵심 자산이다.**
- **태양전지 Solar MissionLoop는 1차 적용 사례이며, K-문샷 12대 미션 공통 AI로 확장된다.**

용어 설명 (Glossary of Terms)

용어	의미
Agentic AI	목표를 받아 여러 단계의 작업을 계획·실행하는 AI(문헌조사·실험설계·도구호출·보고서 작성).
Control Plane	여러 시스템을 직접 대체하지 않고 연결·제어하는 계층. 새결AI의 핵심.
Routing	요청을 적절한 모델·도구·실행환경으로 보내는 결정.
Mask & Hydrate	민감정보를 가린 뒤 외부 AI를 쓰고 내부에서 원문을 복원하는 방식.
TEE	Trusted Execution Environment. 클라우드·서버 안에서도 격리된 보안 영역에서 실행.
MissionLoop	탐색·실험·분석·수정을 반복하는 폐루프 연구 운영 단위.
Tree search / Pruning	가능성을 트리로 펼쳐 탐색하고, 유망하지 않은 가지를 조기에 제거하는 방법.
Bayesian optimization	적은 실험으로 좋은 조건을 찾는 최적화. 후보물질·공정조건·파라미터에 적용.
Privacy-constrained acquisition	성능 향상과 정보 누출 위험을 함께 고려해 다음 실험·실행 경로를 선택하는 기준.
Run Ledger	실행·데이터·모델·판단 기록 원장. 재현성·감사·IP 증거 보존.
독자 파운데이션 모델	국내 / 기관이 통제 가능한 기반 AI 모델. 데이터 주권·기술 자립의 기반.
구조화 과학 산출물	조성·공정조건·측정곡선·후보 ranking 등 정해진 필드를 가진 보호 대상 데이터.
PCE / ETL / HTL	광전변환효율 / 전자·정공 수송층. 태양전지 성능·구조 용어.
DFT / HPC	밀도범함수이론 계산 / 고성능 컴퓨팅.

참고문헌 (References)

[1] J. Zhan, H. Shen, Z. Lin, T. He, “PRISM: Privacy-Aware Routing for Adaptive Cloud--Edge LLM Inference via Semantic Sketch Collaboration,” AAAI 2026; arXiv:2511.22788.

[2] A. E. Ghareeb et al., “A multi-agent system for automating scientific discovery (Robin),” Nature, 2026, doi:10.1038/s41586-026-10652-y.

[3] D. Gao, S. Lu, C. Zhang et al., “Autonomous closed-loop framework for reproducible perovskite solar cells,” Nature, 2026, doi:10.1038/s41586-026-10482-y.

[4] C. Lu et al., “Towards end-to-end automation of AI research,” Nature, 2026, doi:10.1038/s41586-026-10265-5.

[5] J. Gottweis et al., “Accelerating scientific discovery with Co-Scientist,” Nature, 2026, doi:10.1038/s41586-026-10644-y.

[6] Y. Yamada, R. T. Lange, C. Lu et al., “The AI Scientist-v2: Workshop-Level Automated Scientific Discovery via Agentic Tree Search,” arXiv:2504.08066, 2025.

[7] D. A. Boiko, R. MacKnight, B. Kline, G. Gomes, “Autonomous chemical research with large language models,” Nature 624:570--578, 2023, doi:10.1038/s41586-023-06792-0.

[8] B. P. MacLeod et al., “Self-driving laboratory for accelerated discovery of thin-film materials,” Science Advances 6:eaaz8867, 2020.

[9] A. M. Bran, S. Cox, O. Schilter, A. D. White, P. Schwaller, “ChemCrow: Augmenting large language models with chemistry tools,” arXiv:2304.05376, 2023; Nature Machine Intelligence, 2024.

[10] N. J. Szymanski et al., “An autonomous laboratory for the accelerated synthesis of novel materials (A-Lab),” Nature 624:86--91, 2023, doi:10.1038/s41586-023-06734-w.

[11] A. Merchant et al., “Scaling deep learning for materials discovery (GNoME),” Nature 624:80--85, 2023, doi:10.1038/s41586-023-06735-9.

[12] H. B. McMahan et al., “Communication-Efficient Learning of Deep Networks from Decentralized Data,” AISTATS 2017; arXiv:1602.05629.

[13] K. Bonawitz et al., “Practical Secure Aggregation for Federated Learning on User-Held Data,” ACM CCS 2016; arXiv:1611.04482.

[14] M. Abadi et al., “Deep Learning with Differential Privacy,” ACM CCS 2016; arXiv:1607.00133.

[15] P. Vepakomma et al., “Split learning for health: Distributed deep learning without sharing raw patient data,” arXiv:1812.00564, 2018.

[16] M. Chrapek, M. Copik, E. Mettaz, T. Hoefler, “Confidential LLM Inference: Performance and Cost Across CPU and GPU TEEs,” arXiv:2509.18886, 2025.

[17] L. Chen, M. Zaharia, J. Zou, “FrugalGPT: How to Use Large Language Models While Reducing Cost and Improving Performance,” TMLR, 2024.

[18] I. Ong et al., “RouteLLM: Learning to Route LLMs with Preference Data,” ICLR, 2025.

[19] I. Gim et al., “Confidential Prompting: Protecting User Prompts from Cloud LLM Providers,” arXiv:2409.19134, 2024.

[20] L. Moreau, P. Missier, “PROV-DM: The PROV Data Model,” W3C Recommendation, 2013.

[21] NIST, “Artificial Intelligence Risk Management Framework (AI RMF 1.0),” NIST AI 100-1, 2023.

[22] NIST, “AI RMF: Generative Artificial Intelligence Profile,” NIST AI 600-1, 2024.

[23] NIST, “Privacy Framework v1.0,” NIST CSWP, 2020.

[24] S. Rose, O. Borchert, S. Mitchell, S. Connelly, “Zero Trust Architecture,” NIST SP 800-207, 2020.

[25] USPTO, “Inventorship Guidance for AI-Assisted Inventions,” Federal Register, 2024.

[26] European Parliament and Council, “Regulation (EU) 2024/1689 (Artificial Intelligence Act),” Official Journal of the EU, 2024.

[27] J. Kirchenbauer, J. Geiping, Y. Wen, J. Katz, I. Miers, T. Goldstein, “A Watermark for Large Language Models,” arXiv:2301.10226, 2023.

[28] J. Snoek, H. Larochelle, R. P. Adams, “Practical Bayesian Optimization of Machine Learning Algorithms,” NeurIPS 2012, pp. 2951--2959.

[29] S. Yao et al., “Tree of Thoughts: Deliberate Problem Solving with Large Language Models,” NeurIPS, 2023; arXiv:2305.10601.

[30] S. Hao et al., “Reasoning with Language Model is Planning with World Model,” EMNLP, 2023; arXiv:2305.14992.

[31] D. Zhang, S. Zhoubian, Z. Hu, Y. Yue, Y. Dong, J. Tang, “ReST-MCTS*: LLM Self-Training via Process Reward Guided Tree Search,” NeurIPS, 2024; arXiv:2406.03816.

[32] Z. Deng et al., “PlanU: LLM Reasoning through Planning under Uncertainty,” NeurIPS, 2025.

[33] Z. Zheng, Z. Xie, Z. Wang, B. Hooi, “Monte Carlo Tree Search for Comprehensive Exploration in LLM-Based Automatic Heuristic Design,” ICML, 2025.

[34] A. Zhou, K. Yan, M. Shlapentokh-Rothman, H. Wang, Y.-X. Wang, “Language Agent Tree Search Unifies Reasoning, Acting, and Planning,” ICML, 2024; arXiv:2310.04406.

[35] Z. Chen, M. White, R. Mooney, A. Payani, Y. Su, H. Sun, “When is Tree Search Useful for LLM Planning? It Depends on the Discriminator,” ACL, 2024.

[36] T. Liu, N. Astorga, N. Seedat, M. van der Schaar, “Large Language Models to Enhance Bayesian Optimization (LLAMBO),” ICLR, 2024.

[37] M. Akke et al., “When Do LLMs Improve Bayesian Optimization? A Systematic Comparison Across Molecular and Protein Design,” NeurIPS, 2025.

[38] R. Gupta, J. Hartford, B. Liu, “LLMs for Bayesian Optimization in Scientific Domains: Are We There Yet?,” Findings of EMNLP, 2025.

[39] A. Kristiadi et al., “A Sober Look at LLMs for Material Discovery: Are They Actually Good for Bayesian Optimization Over Molecules?,” ICML, 2024.

[40] Y. Zhu et al., “Sample-efficient Multi-objective Molecular Optimization with GFlowNets,” NeurIPS, 2023.

[41] H. W. Sprueill, C. Edwards, M. V. Olarte, U. Sanyal, H. Ji, S. Choudhury, “Monte Carlo Thought Search: LLM Querying for Complex Scientific Reasoning in Catalyst Design,” Findings of EMNLP, 2023.

[42] K. Swanson, G. Liu, D. B. Catacutan, A. Arnold, J. Zou, J. M. Stokes, “Generative AI for designing and validating easily synthesizable and structurally novel antibiotics (SyntheMol),” Nature Machine Intelligence, 2024.

[43] Z. Qi, M. Ma, J. Xu, L. L. Zhang, F. Yang, M. Yang, “Mutual Reasoning Makes Smaller LLMs Stronger Problem-Solvers,” ICLR, 2025.

[44] S.-E. Baek, J.-M. Lee, S.-B. Kim, T.-H. Oh, “A Language-Guided Bayesian Optimization for Efficient LoRA Hyperparameter Search,” ICML, 2026; arXiv:2602.11171.

[45] D. Silver et al., “Mastering the game of Go with deep neural networks and tree search,” Nature 529:484--489, 2016.

[46] D. Silver et al., “A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play,” Science 362:1140--1144, 2018.

[47] A. Fawzi et al., “Discovering faster matrix multiplication algorithms with reinforcement learning (AlphaTensor),” Nature 610:47--53, 2022.

[48] J. Jumper et al., “Highly accurate protein structure prediction with AlphaFold,” Nature 596:583--589, 2021.

[49] H. Jang, S. Nooshabadi, K. Kim, H.-N. Lee, “Circular Sphere Decoding: A Low Complexity Detection for MIMO Systems with General Two-dimensional Signal Constellations,” IEEE Trans. Vehicular Technology 66(3):2085--2098, 2017.

[50] H.-N. Lee et al., “Robust Iterative Tree-Pruning Detection and LDPCC Decoding,” IEEE J. Selected Areas in Communications, 2005.

[51] H.-N. Lee et al., “Performance Analysis on LDPC-Coded Systems over Quasi-Static (MIMO) Fading Channels,” IEEE Trans. Communications, 2008.

[52] J.-H. Kim et al., “Efficient and Stable Perovskite Solar Cells with a High Open-Circuit Voltage Over 1.2 V Achieved by a Dual-Side Passivation Layer,” Advanced Materials, 2022.

[53] J.-H. Kim et al., “Efficient and Stable Quasi-2D Ruddlesden--Popper Perovskite Solar Cells by Tailoring Crystal Orientation and Passivating Surface Defects,” Advanced Materials, 2023.

[54] L. A. Castriotta et al., “Challenges, technological pathways and trade-offs of perovskite solar modules for long-term operation,” Nature Energy, 2026.

[55] P. Sun et al., “From Literature to Lab: Closed-Loop Advancement of Perovskite Solar Cells via Domain Knowledge Guided LLM,” arXiv:2602.04914, 2026.

[56] 과학기술정보통신부 / 대한민국 정책브리핑, “과학기술 12대 국가미션 해결 K-문샷 프로젝트 개시” 및 K-문샷 추진전략, 2026.

[57] GIST INFONET Lab / ITRC 블록체인지능융합센터 / LiberVance(기업부설연구소), “Agentic AI 기반 프라이버시·지식재산 보호형 Self-Driving Lab 연구동향 보고서(내부자료),” 2026.06.09.

[58] GIST, “GIST 주도 K-문샷 공공연구 데이터 주권형 범용 Agentic AI 플랫폼 개발 사업계획 / 새결 Private AI Solar MissionLoop 과제기획 제안서(내부자료),” 2026.

[59] GIST(ChatGPT Deep Research 보조), “AI Scientist의 현재 기술 성숙도(내부자료),” 2026.05.26.

[60] V. Mnih, K. Kavukcuoglu, D. Silver et al., “Human-level control through deep reinforcement learning,” Nature 518(7540):529--533, 2015, doi:10.1038/nature14236.

[61] McKinsey & Company, “Sovereign AI: Building ecosystems for strategic resilience and impact,” 2026 (2025.12 분석) --- 2030년 글로벌 약 $500--600B 규모, AI 지출의 30--40%가 주권 요건의 영향을 받을 것으로 추정.

[62] IBM Security & Ponemon Institute, “Cost of a Data Breach Report 2025,” 2025 --- 글로벌 평균 $4.44M, 미승인 외부 AI(shadow AI) 관련 추가 비용 약 $670K, 조사 기관의 63%가 AI 거버넌스 정책 부재.