

발간등록번호

00-00000000-0000000-00

ISBN 000-00-0000-000-0 0000

지식재산 데이터 큐레이션 연구개발을 위한 선행연구

2026. 07



제 출 문

지식재산처장 귀하

본 보고서를 “지식재산 데이터 큐레이션 연구개발을 위한 선행연구” 과제의 최종 보고서로 제출합니다.

2026년 7월 3일

- 주관연구기관명 : 한국특허전략개발원
- 연구기간 : 2026년 5월 11일 ~ 2026년 7월 3일
- 주관연구책임자 : 권병기 (한국특허전략개발원 연구위원)
- 참여연구원
 - 연구원 : 전상규 (한국특허전략개발원 전문위원)
 - 연구원 : 최진영 (한국특허전략개발원 책임연구원)
 - 연구원 : 황지훈 (한국특허전략개발원 주임연구원)

연구보고서 목차

제 1장 서론	1
제1절 연구의 필요성	1
제2절 연구의 내용 및 방법	7
제2장 환경분석	8
제1절 주요국 정책 동향	8
제2절 시장 동향	13
제3절 기술 동향	14
제3장 R&D 추진전략	27
제1절 추진 방향 도출	27
제2절 비전 및 목표	34
제3절 핵심 추진과제	35
제4절 추진전략	52
제4장 기대효과	54
제1절 기술적 효과	54
제2절 경제적 효과	55
제3절 사회적 효과	56
제 5장 결론	57

표 목차

표 1. 지식재산처 IP-AI 솔루션 비교	6
표 2. 비정형 데이터 처리단계 예시	15
표 3. 문서 처리 방식 비교	17
표 4. AI 기반 R&D 전략	23
표 5. 5일 부트캠프 운영 개요	25
표 6. 3단계 수익 구조	26
표 7. 환경 분석 및 기술개발 가능성 분석	27
표 8. SWOT 분석	31
표 9. SWOT 시사점	32
표 10. 특허문헌 내 비정형 데이터 분석·활용 요건	35
표 11. 컨소시엄 구성 예시	36
표 12. 지식재산 데이터 큐레이션 중점 지원 분야	40
표 13. 반도체/디스플레이 분야 과제 개요	41
표 14. 이차전지 분야 과제 개요	42
표 15. 첨단바이오 분야 과제 개요	44
표 16. ICT표준-차세대 통신 분야 과제 개요	46
표 17. ICT표준-차세대 통신 분야 과제 개요	47
표 18. ICT표준-첨단로봇·제조 분야 과제 개요	49
표 19. 분야별 비정형 데이터셋	50
표 20. 분야별 기술 난제 해결 알고리즘	51



그림 목차



그림 1. 미지원 과제 대비 IP-R&D 지원 과제의 성과 효과	2
그림 2. 노동집약적인 IP-R&D 프로세스	2
그림 3. 실제 IP-R&D 사례(반도체 분야 공급망 해결(포토레지스트 레시피)) ..	5
그림 4. The 2021 MGI Strategic Plan	8
그림 5. USPTO launches new Open Data Portal for easy, quick access to data	9
그림 6. Executive Order 14110 on Artificial Intelligence	10
그림 7. Action Plan for Utilization of AI technology (FY2022-2026 edition)	11
그림 8. 정부상호운용성 프레임워크(GIF) 데이터 모델 구조	12
그림 9. Market Size Forecast 2024–2034	13
그림 10. Scientific Data Management System Market Growth	14
그림 11. 지능형 문서처리 프로세스	16
그림 12. 지능형 문자인식 예시	18
그림 13. 자연어 처리 적용 예시	18
그림 14. 데이터 검증 및 AI모델 학습 흐름도	19
그림 15. 특허 분석 엔진(PACE)	21
그림 16. The architecture of the pipeline to create DeepPatent2	22
그림 17. 시트린-파나소닉 협업 성과	23
그림 18. 시트린 사의 AI 학습용 특허 데이터셋 구축 프로세스	24
그림 19. 시트린파나소닉의 AI 학습용 특허 데이터셋 구축을 통한 유기 반도체 후보 탐색 사례....	24
그림 20. 엘스비어 리액시스 활용 사례	26
그림 21. 비전 및 목표	34
그림 22. 국가전략기술분야의 검증 가능 테마 도출 프로세스	35
그림 23. TRL 예시	37



그림 목차



그림 24. IP-AX 제품 개발 단계	38
그림 25. 우선지원분야 도출 계획	40
그림 26. 반도체 큐레이션 대상 항목 예시	42
그림 27. 이차전지 큐레이션 대상 항목 예시	44
그림 28. 첨단바이오 큐레이션 대상 항목 예시	46
그림 29. 인공지능 큐레이션 대상 항목 예시	48
그림 30. 첨단로봇제조 큐레이션 대상 항목 예시	50

요 약 문

R&D 방식이 기존 경험·이론 중심의 반복 실험 구조에서 데이터·AI 기반 예측 중심 구조로 전환되며 기술 혁신 패러다임이 변화하고 있다. 그러나 현재 R&D 체계는 여전히 경험 기반 의사결정에 의존하고 있으며, 특허데이터 또한 비정형·법률 중심 구조로 존재하여 AI 활용 및 데이터 자산화에 한계가 있다. 전 세계 기술 정보의 약 80%가 담긴 특허는 기술 공백, 경쟁력, 분쟁 위험 등을 예측할 수 있는 국가 핵심 데이터 자산이나, IP-R&D는 기술료 4.4배 등 우수한 성과에도 불구하고 수작업 기반의 노동집약적 구조로 인해 '24년 기준 정부 R&D 응용·개발단계 과제 27,128개 중 약 2%인 580개 과제만 지원되는 등 확산에 한계가 있다. 이에 비정형 지식재산 데이터를 고부가가치 연구데이터로 전환하고, 이를 기반으로 지능형 R&D 체계를 구축하기 위한 데이터·AI 기반 연구개발 추진이 시급하다.

국내에는 방대한 특허 및 연구데이터가 축적되어 있으나 대부분 비정형·과편화된 상태로 존재하여 즉시 활용 가능한 AI 학습·분석용 데이터로 전환되지 못하고 있다. 특히 특허 기반 고품질 AI 학습데이터 구축은 높은 기술 난이도와 비용 부담으로 인해 민간 중심으로는 한계가 있으므로, 민관이 독자적으로 감당하기 어려운 고난도 데이터의 자산화를 공공 부문이 주도하는 선제적 마중물 투자가 요구된다.

본 연구는 주요국 정책·제도·기술 동향과 시장 동향을 종합적으로 조사·분석하여 지식재산 데이터 큐레이션 기반 R&D 분야의 추진 여건을 진단하고, 사업 추진의 기본방향, 전략목표, 핵심 지원 분야, 중점 투자영역 및 기대효과를 도출하였다. 환경분석 결과, 미국은 연방 차원의 데이터 표준화 및 AI-Ready 생태계를 주도하고 있으며, 일본은 관(官) 주도의 데이터 표준화와 업무 효율화 중심으로 AI를 도입하는 등 주요국은 지식자산이 AI 학습에 즉시 활용될 수 있도록 데이터 표준화, 특허 데이터 개방·제공 체계 고도화, AI 기반 분석 인프라 확충을 가속화하고 있는 것으로 나타났다. 또한 버티컬 AI와 과학 데이터 관리 시스템(SDMS) 시장이 급성장하면서 고품질 데이터 확보가 AI 성패를 좌우하는 데이터 중심 AI(Data-Centric AI) 시대로 진입한 것으로 분석되었다.

환경분석과 SWOT 분석을 토대로 본 연구는 사업 비전을 ‘비정형 IP 데이터의 지능적 자산화를 통한 R&D 패러다임 대전환 및 국가 전략기술 초격차 선점’으로 설정하였다. 목표는 ‘비정형 IP 데이터의 고정밀 구조화 및 표준화를 통한 고부가가치 연구 데이터 자산 체계 확립’과 ‘IP-AI 기반 처방형 솔루션 구축을 통한 초격차 기술 확보 및 R&D 메커니즘 혁신’으로 제시하였다. 추진전략은 ① 전략적 IP 데이터 자산화 체계 확립, ② AI 기반 지능형 R&D 혁신 플랫폼 고도화, ③ 기술 가치사슬 확장 및 실증 가속화로 구성하고, 큐레이션 기반 PoC 검증, 데이터 레이크 구축, 기술군 클러스터링, 수요기업 연계 실증 및 AI 모델 고도화를 통해 전주기 R&D 선순환 체계를 구축하는 방향으로 설계하였다.

세부적으로는 비정형 특허 데이터를 마쿠쉬 구조식, 실험수치, 도면 등 핵심 정보를 포함한 AI-Ready 데이터로 정밀 큐레이션하여 국가 지식자산 체계를 구축하고자 한다. 중점 지원 대상은 12대 국가전략기술 중 비정형 데이터 비중이 높고 특허 기반 R&D 수요가 집중되는 반도체·디스플레이, 이차전지, 첨단바이오, 차세대통신, 인공지능, 첨단로봇·제조 등 데이터 기반 R&D 효과가 큰 영역으로 설정하였다. 지원 과제는 IP 데이터를 R&D의 결과물이 아닌 핵심 생산·투입요소로 활용하고, AI 학습용 고품질 데이터셋 구축 및 IP-AI 기반 지능형 R&D 솔루션 도출이 가능한 과제로 설정하였다. 또한 산·학·연 수요기관, 데이터 큐레이션 기관, AI·SW 전문기업이 참여하는 개방형 컨소시엄을 중심으로 연구 현장 실증 가능성을 갖춘 과제를 지원하는 체계로 기획하였다.

단계별 지원 체계는 1단계에서 연구데이터와 특허·논문 등 외부 데이터를 정제하여 AI 학습용 고품질 데이터셋을 구축하고 데이터 자산화를 수행하며, 2단계에서는 AI 기반 연구모델과 IP-AI 처방형 R&D 솔루션을 활용하여 연구설계, 가상실험, 신규 아이디어 도출, 산업 현장 실증을 수행하고 이를 기술 확보, 특허화, 제품 개발로 연계하는 구조로 마련하였다. 이를 통해 비정형 IP 데이터 큐레이션, AI 학습데이터 구축, 지능형

솔루션 개발, 산업 현장 실증으로 이어지는 전주기 지원 체계를 구축하고자 하였다.

본 연구가 도출한 사업은 기술적으로 고난도 기술 문헌 분석 및 정보 추출 분야의 독자적 기술력을 확보하고, 데이터 수집부터 지능형 분석까지 연결되는 기술 정보 선순환 체계를 구축함으로써 국가 차원의 기술 주권 확립에 기여할 것으로 기대된다. 경제적으로는 중복 투자 피해 경감, 연구개발 자원의 효율적 배분, 연구 전주기 생산성 제고 및 신개념 지식 서비스 산업 활성화가 기대된다. 사회적으로는 AI 기반 처방형 솔루션의 연구 현장 적용을 통해 R&D 시행착오를 단축하고, 연구자 간 정보 격차 해소와 공정한 기술 거래·활용 촉진에 기여할 수 있다.

결과적으로 본 연구는 지식재산 데이터 큐레이션 기반 R&D 추진의 객관적 타당성과 세부 추진과제를 도출함으로써, 경험 중심의 R&D에서 데이터·AI 기반 지능형 연구개발로 전환하기 위한 정책적·기술적 기반을 제시하였다. 제안된 사업이 추진될 경우, IP 데이터가 핵심 R&D 자원으로 전환되어 국가 R&D 생산성이 제고되고, 기술과 데이터 융합 기반의 새로운 비즈니스 생태계 확장과 국가 지식재산 총역량 강화에 기여할 것으로 기대된다.

S U M M A R Y

The R&D paradigm is shifting from a conventional structure centered on experience- and theory-based iterative experimentation to a data- and AI-driven structure focused on prediction, thereby transforming the broader paradigm of technological innovation. However, the current R&D system still relies heavily on experience-based decision-making, while patent data remains largely unstructured and legal-document-oriented, limiting its use for AI applications and its transformation into strategic data assets. Patents, which contain approximately 80% of the world's technical information, are core national data assets that can be used to predict technological gaps, competitiveness, and dispute risks. Nevertheless, although IP-R&D has demonstrated strong performance, including a 4.4-fold increase in technology royalties, its expansion has been constrained by a labor-intensive, manual operating structure. As of 2024, only 580 projects, or approximately 2%, of the 27,128 government R&D projects in the applied and development stages received IP-R&D support. Accordingly, it is urgent to promote data- and AI-based R&D that converts unstructured intellectual property data into high-value research data and establishes an intelligent R&D system based on such data.

Although Korea has accumulated vast amounts of patent and research data, most of it exists in an unstructured and fragmented form and has not yet been converted into data that can be immediately used for AI training and analysis. In particular, the construction of high-quality AI training data based on patents involves high technical complexity and significant costs, making it difficult to be led by the private sector alone. Therefore, proactive public-sector investment is required as a catalyst to support the assetization of highly complex data that the private sector cannot independently undertake.

This study comprehensively investigated and analyzed policy, institutional, technological, and market trends in major countries to diagnose the conditions for promoting R&D based on intellectual property data curation. It also derived the basic direction, strategic goals, key support areas, priority investment domains, and expected effects of the proposed initiative. The environmental analysis found that major countries are accelerating efforts to standardize data, advance patent data opening and provision systems, and expand AI-based analytical infrastructure so that knowledge assets can be immediately utilized for AI training. The United States is leading the development of a federally driven data standardization and AI-ready ecosystem, while Japan is introducing AI under a government-led approach focused on data standardization and operational efficiency. In addition, the rapid growth of vertical AI and scientific data management system markets indicates that the world has entered the era of Data-Centric AI, in which the securing of high-quality data determines the success of AI.

Based on the environmental and SWOT analyses, this study established the business vision as “a fundamental transformation of the R&D paradigm and the securing of national strategic technology super-gap competitiveness through the intelligent assetization of unstructured IP data.” The goals were defined as “establishing a high-value research data asset system through the high-precision structuring and standardization of unstructured IP data” and “securing super-gap technologies and innovating the R&D mechanism through the development of IP-AI-based prescriptive solutions.” The promotion strategies consist of: 1) establishing a strategic IP data assetization system; 2) ad-

vancing an AI-based intelligent R&D innovation platform; and 3) expanding the technology value chain and accelerating demonstration. The initiative was designed to build a virtuous cycle across the entire R&D process through curation-based PoC verification, data lake construction, technology-cluster analysis, demand-company-linked demonstration, and AI model advancement.

Specifically, the initiative aims to establish a national knowledge asset system by precisely curating unstructured patent data into AI-ready data that includes key information such as Markush structural formulas, experimental values, and drawings. The priority support areas were selected from among the 12 national strategic technologies, focusing on fields with a high proportion of unstructured data and strong demand for patent-based R&D. These include semiconductors and displays, secondary batteries, advanced bio, next-generation communications, artificial intelligence, and advanced robotics and manufacturing, all of which are expected to generate significant effects from data-driven R&D. Supported projects were designed to use IP data not as an output of R&D, but as a core production and input factor, and to enable the construction of high-quality AI training datasets and the development of IP-AI-based intelligent R&D solutions. In addition, the support system was planned around open consortiums involving industry, academia, research demand organizations, data curation institutions, and AI/SW-specialized companies, with priority given to projects that have strong potential for field demonstration in actual research settings.

The phased support system was structured so that, in the first stage, research data and external data such as patents and papers are refined to build high-quality datasets for AI training and to carry out data assetization. In the second stage, AI-based research models and IP-AI prescriptive R&D solutions are applied to research design, virtual experimentation, new idea generation, and industrial field demonstration, which are then linked to technology acquisition, patenting, and product development. Through this structure, the initiative seeks to establish an end-to-end support system connecting unstructured IP data curation, AI training dataset construction, intelligent solution development, and industrial field demonstration.

The initiative derived from this study is expected to contribute technologically to securing independent capabilities in the analysis of highly complex technical documents and information extraction, while establishing a virtuous cycle of technical information that connects data collection with intelligent analysis. This will ultimately support the establishment of national technological sovereignty. Economically, the initiative is expected to reduce losses from duplicate investment, improve the efficient allocation of R&D resources, enhance productivity across the entire research cycle, and stimulate the growth of new knowledge service industries. Socially, the application of AI-based prescriptive solutions in research settings can shorten R&D trial-and-error processes, reduce information gaps among researchers, and promote fair technology transactions and utilization.

In conclusion, this study presented the policy and technological foundation for transitioning from experience-based R&D to data- and AI-driven intelligent R&D by deriving the objective feasibility and detailed implementation tasks for promoting R&D based on intellectual property data curation. If the proposed initiative is implemented, IP data will be transformed into a core R&D resource, thereby enhancing national R&D productivity, expanding a new business ecosystem based on the convergence of technology and data, and strengthening the nation's overall intellectual property capabilities.

제 1장 서론

제1절 연구의 필요성

- (R&D 패러다임 전환) R&D 방식이 기존 경험·이론 중심에서 데이터·AI 주도형으로 급격히 전환되며 기술 혁신의 뉴노멀(New Normal)로 부상
 - 기존 연구자의 직관과 반복 실험(Trial & Error)에 의존하던 방식은 이제 기술 난도 및 비용 상승으로 인해 한계 봉착
- (AI 기반 R&D 성과) 글로벌 기업은 AI를 활용해 연구 기간을 획기적으로 단축하고 있으며, 주요국은 기술 패권 경쟁 시대 속에 미래 시장 선점을 위한 패권형 지식재산 창출에 사활을 거는 추세
 - 글로벌 선도 기업은 데이터 자산화를 통해 장기간 소요되는 연구를 초단기로 단축하는 파괴적 혁신 사례 창출
 - (구글 딥마인드) AI 모델 'GNoME'을 통해 800년 분량의 신소재 결정 구조 220만 개를 단 수주 만에 발견(Nature, '23.11)
 - (시트린 인포매틱스) 특허·논문의 텍스트를 AI로 구조화(Curation)하여 파나소닉 등의 신소재 개발 기간을 50% 이상 단축(美 Citrine, '23)
- (특허 데이터 가치 및 IP 전주기 혁신) 전 세계 기술 정보의 약 80%가 담긴 특허는 독점 권리 정보뿐만 아니라 논문·시장 자료가 제시하지 못하는 기술 공백, 경쟁력, 분쟁 위험 등을 예측할 수 있는 전 세계 유일의 국가 핵심 데이터 자산임
 - (전략적 가치) 독점 기술 정보로서 전세계 6억 건 이상의 특허 데이터는 최적의 R&D 방향 설정과 고효율 R&D 수행의 원재료로 활용이 가능한 핵심 자료이자, 국가 지식재산 총역량을 강화하는 국가 핵심 전략 자산
 - (활용 예) 원천·경쟁사 특허의 권리 소멸 시점 분석을 통해 법적 리스크 없는 기술 도입 및 개량 연구 전략 수립 가능
 - (IP기반 혁신지원) 글로벌 특허 빅데이터와 AI 융복합을 통해 IP 전주기(창출-활용-보호-기반)에 걸친 혁신기술 개발을 체계적으로 지원함으로써 국가 R&D 효율성을 제고하고 경제 안보를 강화할 필요성 증대
 - 신규 기술 정보의 80%가 특허를 통해서만 공개, 미활용 시 심각한 R&D 중복 투자가 발생할 수 있음을 경고(출처: WIPO)
- (AI-Ready Data 확보 시급) R&D 패러다임이 AI 중심으로 전환됨에 따라 생산성 혁신의 관건은 자본·인력이 아닌 AI 역량에 있으며, 이를 실현하기 위해서는 국가 R&D 성과물을 기계가 즉시 학습할 수 있는 고품질의 AI-Ready Data로 구축·확보하는 것이 무엇보다 시급
 - (AI 중심 기술 패권) R&D 성과 의존도가 자본·인력에서 AI 활용 역량으로 이동함에 따라 미래 기술 격차를 좌우할 R&D 혁신이 시급
 - (중국 사례) 텐센트는 '26년 3월 기존 연구소 해체·박사급 인력을 대거 해고하고, 자체 파운데이션 모델 고도화에 R&D 자본과 역량을 집중 투입하는 등 AI 중심의 연구체제로 전환
 - (AI-Ready Data 필수) AI 파운데이션 모델도 중요하나, AI 성능은 결국 학습하는 데이터가 좌우하므로 고품질 데이터 확보가 필수

- (AI-Ready data) AI 모델이 전처리 없이 즉각 학습·추론에 활용할 수 있도록 기계 가독성 (Machine-Readable), 품질, 맥락정보가 최적화된 상태의 데이터를 의미
- “AI가 R&D에 투입되면 최대 10배 R&D 생산성의 시대가 가능하나 AI가 쓸수 있는 데이터가 없다. 수많은 국가 R&D 성과는 라벨링이 없어 쓸 수가 없다. 바이오 분야 혁신을 위해서는 AI-Ready 데이터가 있어야 한다” (국가바이오혁신위원회('26.4.16) 중 김태영 위원)

□ (특허데이터를 활용한 R&D 효율화 한계) IP-R&D는 기술료 4.4배 등 성과가 탁월하나, 수작업 기반의 노동집약적인 구조로 인해 정부 과제 지원율이 상대적으로 낮으며, R&D 효율의 전면적 확산을 위해 분석 프로세스 혁신 및 데이터 활용 기반의 체계 개선이 시급함

- (IP-R&D 지원) 지식재산처는 특허분석을 기반으로 국가 R&D의 기술 난제를 해결하고, 핵심특허를 확보하기 위한 R&D 전략(IP-R&D)을 지속 지원 중('09~), IP-R&D를 통해 우수특허 확보, 중복투자 방지뿐만 아니라 비용 및 시간 절감 등 효율적인 R&D 수행이 가능
- 또한, IP-R&D를 지원받은 과제는 일반 R&D 과제 대비 우수특허·해외특허 등 질적 특허성과는 1.3~2.1배, 특허 이전율은 1.3배, 기술이전 계약당 기술료는 4.4배 높음('19~'23년 성과)

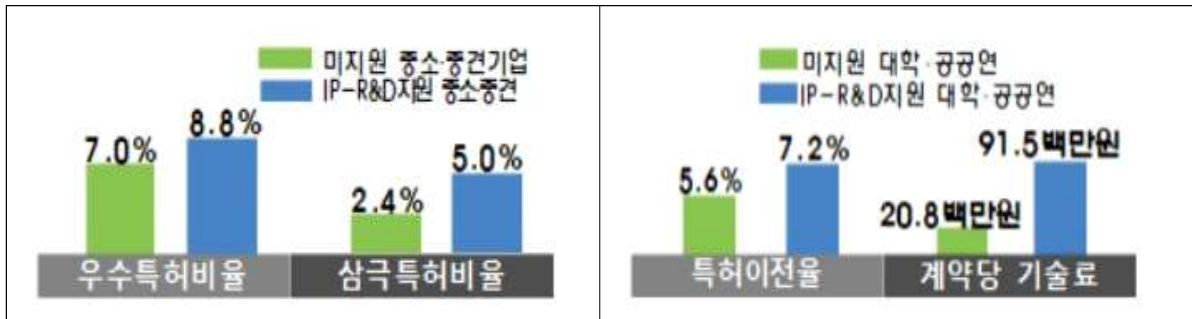


그림 1. 미지원 과제 대비 IP-R&D 지원 과제의 성과 효과

- (노동집약적 수행 프로세스 개선필요) IP-R&D 지원으로 R&D 성과는 크게 향상되었으나, 노동집약적인 수행 방법과 한정된 예산으로 인해 국가 R&D 전반으로의 확산은 부족
- '24년 기준 한정된 예산으로 인해 정부 R&D 응용·개발단계 과제(27,128개) 중 IP-R&D를 지원한 과제 수는 약 2%(580개)에 불과
- 현행 IP-R&D 프로세스는 방대한 특허 데이터를 수작업으로 일일이 검토하고 데이터화하는 방식으로 많은 인건비 및 시간 소요의 한계가 발생

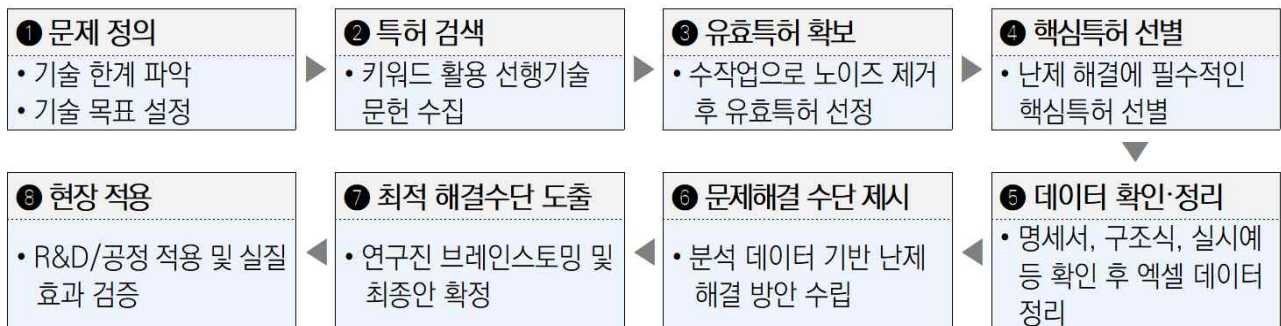


그림 2. 노동집약적인 IP-R&D 프로세스

□ (개념) 특허 데이터 지능화(Patent Data Intelligence)는 특허문헌 내 비정형 실험 데이터를 기계 독해가 가능한 형태로 자산화하여, 연구개발 전 과정을 데이터 기반으로 혁신하는 최상위 지능형 단계

- 특허 큐레이션¹⁾과 지능형 R&D²⁾는 데이터 기반 기술 주권 확보를 위해 상호 유기적으로 결합되어야 할 핵심 요소

□ (데이터 활용 장벽) 특허 데이터는 AI 기반 R&D 혁신의 핵심 자산이나, 특허 고유의 복합적 언어 장벽과 데이터 파편화로 인해 실제 R&D 현장 활용률 저조

- (언어 장벽) 기술 내용을 법률적 용어로 방어하는 특수한 문체로 인해 일반적인 자연어 처리(NLP) 기술로는 핵심 정보(목적, 구성, 효과 등) 추출 및 해석에 한계
- (비정형성) R&D 필수 정보인 실험 수치, 화학 구조식, 도면 등이 비정형 이미지나 텍스트로 산재해 있어 AI가 직접 학습하기 어려운 구조적 난제 존재
 - (세계경제포럼, WEF) '26년경 고품질 AI 학습용 텍스트 데이터가 고갈될 것이라 경고하며, 유일한 대안으로 특허와 같은 전문 데이터의 활용을 제시하였으나, 현재 화학·소재 분야의 핵심 데이터(실험수치, 화학식) 대부분이 비정형(표·이미지)으로 잠겨 있어 활용 불가능한 데이터 림보 상태에 직면
 - 또한, 다수의 특허에는 출원인 등 정형 데이터뿐만 아니라 화학 구조식, 염기서열, 도면 등 고난도의 비정형 정보가 파편화되어 있으며, R&D에 AI 도입을 위해서는 비정형 특허 정보를 AI가 즉각 학습할 수 있도록 표준 데이터로 사전 구축(큐레이션)하는 작업이 필요
- (IP 데이터 자산화 인프라) 기존 인프라가 단순 검색 위주에 머물러 생성형 AI 학습 및 기술 추론 수요에 대응하지 못하는 가운데, 정부의 데이터 기반 R&D 강조에도 불구하고 실제 연구 현장의 IP 데이터를 자산화하여 즉시 활용할 수 있는 인프라는 여전히 미비한 실정임
 - '지식재산정보 검색 서비스(KIPRIS)' 등 기존 특허 DB는 키워드 매칭 중심의 '단순 검색'에 최적화되어, AI의 인과관계 학습 및 추론 데이터로는 활용 불가. 또한, 개별 기업이 감당하기 어려운 고품질 학습 데이터 부족 현상이 심화됨에 따라 국가 차원의 큐레이션 지원 절실
 - '22년 기준 정부 R&D 과제(76,052개) 중 응용·개발단계 과제(35,457개)에 대하여 '특허로 R&D'가 지원된 과제 수는 약 2%(연간 약 700개)에 불과. R&D의 ①적시성·신속성 확보, ②전략성·효율성 강화, ③성과활용 고도화 등 혁신을 뒷받침하기 위해 R&D 전주기 연계 특허 빅데이터 활용 확대 필요('24.12., 국가지식재산위원회)
- 따라서, 민간이 독자적으로 감당하기 어려운 고난도 데이터의 자산화를 공공 부문이 주도하여, 특허 기반 고품질 AI 학습데이터가 연구 현장에 끊임없이 공급되고 다양한 혁신 서비스가 창출되는 데이터-AI 선순환 생태계 조성 마련이 필요

□ (AI 데이터로서 특허의 가치) 특허는 공공·민간 R&D를 포괄하며 표준화·완결성을 갖춘 AI 학습 최적화 데이터로 논문 대비 빠른 공개와 상세 정보 수록으로 최신 기술 트렌드를 즉각 데이터화할 수 있는 독보적 정보원

1) (특허 큐레이션) 특허 명세서 내 비정형 데이터(텍스트, 표, 도면 등)를 추출·정제하여 기계가 즉시 학습 및 처리 가능한 구조적 정보로 변환하는 활동

2) (데이터 기반 지능형 R&D) 큐레이션된 데이터 및 인공지능 기술을 활용하여 신물질 발굴, 공정 최적화, 표준특허 매핑 등 연구개발의 의사결정을 가속화하고 초격차 가치를 창출하는 활동

○ **(공공·민간 풀데이터셋)** 특허는 공공 데이터(국가 R&D 성과의 결과물) 및 민간 데이터를 모두 포함한 가장 광범위한 R&D 실측 데이터

- 동 사업의 R&D 전문기관인 한국특허전략개발원은 국가 R&D 11대 연구성과(논문, 표준 등) 중 특허 성과 유통관리 전담기관으로 지정되어(과기부) 특허성과 관리 중

- 논문은 학계 중심으로 출판되는 한편, 특허는 민간이 독점권을 위해 스스로 공개한 데이터로, 기업의 내부문서(영업비밀) 이외에는 외부에서 접근 가능한 유일한 민간 R&D 정보

○ **(AI 활용 최적화 문서)** 특허는 국가정책적으로 표준화된 문서이고, 특허심사를 위한 법적요건에 의해 문서 내 완결도 높은 문서

- 타 문헌의 참조로 끝내는 논문과 달리 특허는 특허법 제42조제3항에 따라 통상의 기술자가 문서만 읽고도 실시할 수 있도록 실험 과정 등이 처음부터 끝까지 문서 내에 기재됨

- 또한, 전체 특허 중 26%('25년 기준)이 우선심사 신청에 따라 빨리 등록되어 공개되고, 출원일 인정을 위해 논문보다 특허를 먼저 출원하며, 논문이 리뷰 과정, 저널 변경 등에 따라 상당시간(1년이상) 소요되는 점 등을 고려하였을 때, 특허는 데이터화 시킬 수 있는 가장 최신의 연구개발 데이터로 판단할 수 있음

□ **(IP-R&D의 한계 탈피 및 지능형 R&D 내재화 가능)** 특허 데이터 큐레이션 기반의 AI 학습 데이터를 활용하여 지능형 솔루션을 개발함으로써, 현행 IP-R&D의 노동집약적 한계점을 근본적 해소하고 R&D 혁신 가능성을 제고

○ **(IP-R&D의 한계)** IP-R&D는 압도적인 성과(일반과제 대비 특허창출 2.1배 기술료 수입 4.4배 차이 발생)와 법적 실시 의무(「국가전략기술육성법」 등에 따라 중요 과제는 의무적 실시)에도 불구하고, 과도한 소요 기간(평균 8개월)과 높은 비용 문제(약 1.2~1.5억원 소요)로 인해 국가 R&D 전반으로 확산하는 데 근본적인 한계가 있음

○ **(IP-R&D 자동 내재화)** 특허 데이터(권리+기술정보)를 기반으로 하므로 특허 데이터를 표준화 및 구조화하여 자산화하는 데이터 큐레이션을 통해 IP-AI 솔루션을 개발할 경우, 핵심특허 확보 및 회피설계 등 IP-R&D의 전문 역량을 R&D 현장에 즉각적으로 내재화할 수 있음

- 또한, 데이터 기반 성공사례(특허문헌에 공개된 제조공정-물성, 성분-기능-물성 등을 연계분석한 레시피(조성물) DB를 구축·분석하여 국내 기업에게 R&D 전략을 제시함으로써 소재 국산화에 성공)를 참고할 때, 큐레이션 기반 IP-AI 솔루션 활용을 통한 IP-R&D 프로세스 개선은 효율적인 R&D 지원체계로서 그 활용 가능성이 매우 높음

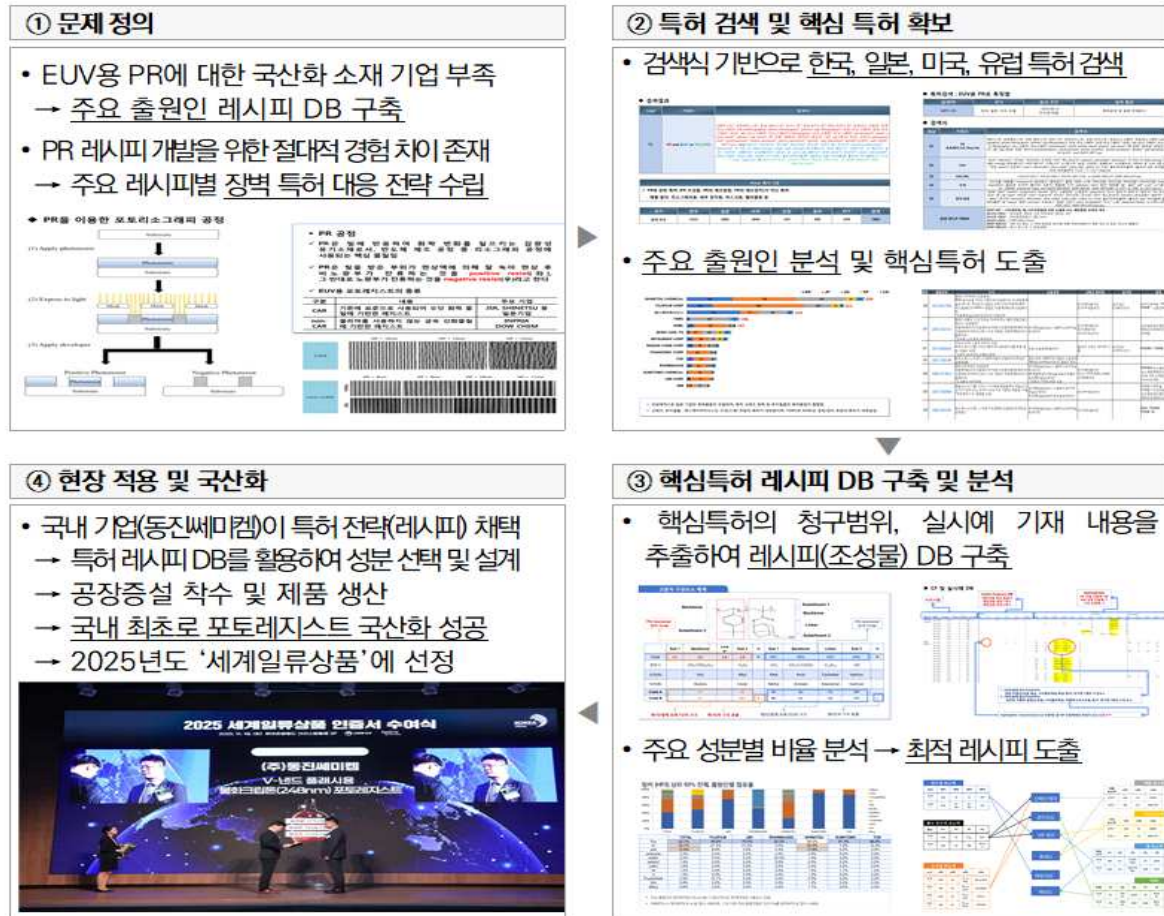


그림 3. 실제 IP-R&D 사례(반도체 분야 공급망 해결(포토레지스트 레시피))

- (현재 비정형 데이터 사업 구조) AI 산업 생태계 조성을 위한 범용 데이터셋의 대량 구축 및 산업 현장의 특정 수요에 따른 부분적 비정형 데이터 활용 지원
- (과학기술정보통신부) AI 학습 주도권 확보를 위해 자연어, 영상 등 비교적 정형화가 용이한 비정형 데이터의 대규모 라벨링 및 공공 서비스 혁신용 데이터셋 구축 지원
 - 인공지능 학습용 데이터 구축 사업, 공공선도형 AI 혁신(AX) 프로젝트 등
 - (산업통상자원부 등) 산업군별 현장 문제해결 및 생산성 혁신을 위해 운영 과정에서 발생하는 로그·센서 데이터와 파편화된 기술 정보의 부분적 수집·실증 지원
 - 산업 AI 솔루션 실증·확산 지원, 분야별 데이터 스페이스 구축 사업 등
 - (KISTI) 국내외 논문, 특허 등 대규모의 과학기술 데이터를 AI로 분석하여 연구개발 생산성을 올리는 AI-Review(Science ON) 서비스 제공
 - 문장기반 학습을 통해 기술 분류 자동화, 노이즈 특허 분류, 기술동향 및 유망기술 예측 서비스 등 분석 효율성 제고
 - (차별성 및 타당성) 이러한 부처별 플랫폼 중심의 지원 체계에도 불구하고 연구 현장에서 직면하는 원천 기술 설계 및 제조 레시피 확보 등 근본적인 기술 난제 해결에는 한계가 존재하기에 특허 큐레이션 기반 지능형 솔루션은 특정 산업을 관리하는 운영 플랫폼 구축이 아닌, 연구 현장의 기술 난제를 즉시 해결하는 R&D 최적화 도구 개발이 핵심임

- (연구 현장 밀착형 기술 해결 프로그램 제공) 타 부처 사업이 제조·임상 환경의 운영 및 모니터링에 집중하는 것과 달리, IP-AI 솔루션은 기업이 R&D 설계 단계에서 직면하는 기술적 병목 구간을 특허 내 비정형 지식으로 해결할 수 있는 지능형 솔루션을 제공
- (특허 데이터의 고유성 및 전문성 기반 큐레이션) 반도체, 바이오 등 국가 전략기술의 핵심 설계도와 실험 레시피는 특허 문헌에만 존재하는 고유 자산이며, 이를 텍스트 뿐만 아니라 이미지 등 비정형 데이터도 정밀 큐레이션하고 권리 관계를 해석하여 기술적 해법을 도출하는 것은 지능형 IP 분석 인프라를 보유한 지식재산처의 고유 업무 영역임
- (부처 간 협력 강화 역할) 특허 큐레이션 기반 지능형 솔루션을 통해 도출된 성공 확률이 높은 기술 해결 방안은 산업부나 복지부가 관리하는 산업별 플랫폼에 탑재되어, 실제 제품화(TRL 4~8) 단계의 성공률을 획기적으로 높이는 핵심 동력 역할을 수행하게 됨

표 1. 지식재산처 IP-AI 솔루션 비교

구분	과기정통부 (범용 AI)	산업부/복지부 (현장/운영)	지식재산처 (IP-AI 솔루션)
데이터 핵심	자연어, 영상 등 일반 비정형	로그·센서, 임상·EHR 데이터	특허 명세서 내 비정형 기술 정보
수행 주체	클라우드 소싱, 일반 가공사	개별 기업 현장, 병원 등	지재처 전문 분석 기관 및 전문가
사업 목적	AI 학습용 범용 데이터 구축	현장 공정 최적화 및 운영 지원	연구 개발(R&D) 기술 난제 해결
최종 형태	공공 서비스용 범용 모델	산업/의료 현장 관리 플랫폼	R&D 설계·실증용 지능형 솔루션

- (필요성) 국가 전략기술 주권 확보와 R&D 생산성 극대화를 위해, 특허 문헌 내 복합 비정형 지식의 큐레이션과 IP-AI 지능형 솔루션을 연계한 전 주기적 R&D 패러다임 전환 지원이 필요
- (법적 기반) R&D 현장의 비정형 실험 데이터는 체계적인 관리 부재로 확보가 어렵고, 이를 특허 등 산업재산 정보와 결합하여 AI로 분석·활용하기 위한 제도적 지침이 미비하여 기업의 중복 투자 및 기술 혁신의 한계 지속
 - 개별 연구실 단위로 산재된 비정형 실험값(데이터)은 규격화 및 공유가 어려워 사장되는 경우가 많고, 이를 공공·민간 특허 빅데이터와 융합하여 분석할 수 있는 인프라도 부족하여 기술 난제 해결에 한계 노출
 - 「산업재산 정보의 관리 및 활용 촉진에 관한 법률」 제16조 및 제23조 등을 통해 비정형 데이터를 포함한 산업재산 정보를 R&D에 효과적으로 활용하고, 지능형 정보서비스를 상용화할 수 있는 법적 근거가 명확히 제시됨
 - 산재된 실험 데이터의 자산화와 AI 기반 지식재산 전주기 혁신을 뒷받침할 제도적 기틀이 마련되어, 데이터 주도형 R&D 생태계 조성을 위한 정부의 선제적 마중물 투자가 시급함

제2절 연구의 내용 및 방법

□ 지식재산 데이터 큐레이션 R&D 분야 환경 분석

- 주요국 정책·제도·기술 동향, 시장 동향 등을 종합적으로 조사·분석하여 지식재산 데이터 큐레이션 기반 R&D 분야의 추진 여건 및 환경을 진단
- 특허 데이터 개방 및 AI 기술 적용과 관련된 국내외 우수사례와 지원체계를 분석하여 사업 추진의 시사점과 성공요인을 도출

□ 지식재산 데이터 큐레이션 R&D 추진전략 수립

- 국내외 정책·시장·기술 환경 분석 결과를 토대로 사업 추진의 기본방향과 전략목표를 설정하고, 중장기 관점의 사업 추진전략을 수립
- 신규 R&D 사업 추진 가능성을 고려하여 핵심 지원 분야 및 중점 투자영역을 도출하고 분야별 추진전략을 마련

□ 지식재산 데이터 큐레이션 R&D 기대효과 도출

- 지식재산 데이터 큐레이션 기반 R&D를 통해 창출 가능한 중점 연구개발 과제 및 핵심기술을 도출하고, 분야별 전략적 투자 필요성을 제시
- 연구개발 성과의 기술적 우수성, 사업화 가능성 및 사회·경제적 파급효과를 분석하여 국민 체감 성과 창출 및 국가 경쟁력 제고 측면의 기대효과를 도출

제2장 환경분석

제1절 주요국 정책 동향

(미국) 연방 차원의 데이터 표준화 및 'AI-Ready' 생태계 주도

□ 소재 게놈 이니셔티브(MGI) 기반의 데이터 인프라 고도화

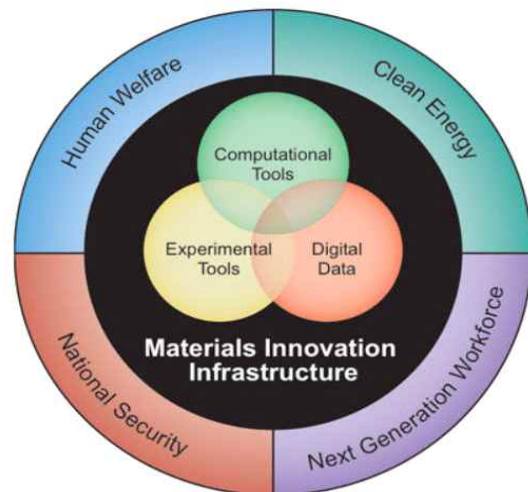
- (개요) 미국 국립표준기술연구소(NIST)는 2011년도부터 추진해 온 '소재 게놈 이니셔티브(MGI; Materials Genome Initiative)'를 통해 신소재 개발 기간과 비용을 절반으로 감축하기 위한 범국가적 데이터 인프라 구축을 총괄
- (전략목표) 2021년 발표된 MGI 전략계획은 소재 데이터 활용 능력(Harnessing the Power of Materials Data)을 핵심 목표로 설정하고, 실험-계산-데이터가 유기적으로 통합된 '소재 혁신 인프라(MII)' 구축에 주력

THE MATERIALS GENOME INITIATIVE AT NIST

The 2021 MGI Strategic Plan

The Materials Genome Initiative marks its first decade with a new strategic plan. Read [the summary](#), or get [the full report](#) (PDF).

The missions of the Materials Genome Initiative and National Institute of Standards and Technology (NIST) are tightly aligned. NIST promotes U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life. MGI addresses precisely these mission elements by providing the means to reduce the cost and development time of materials discovery, optimization, and deployment. Both missions are driven by industrial competitiveness, with the creation of an innovation infrastructure as the means to this end.



출처: Materials Genome Initiative | NIST

그림 4. The 2021 MGI Strategic Plan

○ (핵심 추진 과제) AI-Ready 데이터 구축

- (AI 준비 강화(AI-Readiness)) NIST는 연방 기관 및 연구소가 보유한 소재 데이터가 AI 알고리즘에 즉시 활용될 수 있도록, 데이터의 수집 단계부터 표준화된 메타데이터(Metadata)와 구조(Schema)를 적용할 것을 의무화
(백악관 OSTP Desirable Characteristics of Data Repositories, 2022)
- (FAIR 원칙 구현) 데이터의 검색 가능성(Findable), 접근 가능성(Accessible), 상호 운용성(Interoperability), 재사용성(Reusable)을 보장하는 FAIR 원칙을 MGI 프로젝트 전반에 적용하여 데이터 사일로(Silo) 현상을 방지

○ 주요 플랫폼 및 도구

- (Materials Data Curation System(MDCS)) NIST는 연구자가 비정형 실험 데이터를 손쉽게 구조화된 데이터(XML, JSON 등)로 변환하여 저장·공유할 수 있도록 지원하는 큐레이션 시스템 'MDCS3'를 개발하여 오픈소스로 배포
- (Materials Resource Registry(MRR)) 전 세계에 산재한 소재 데이터 저장소(Repository)와 알고리즘 자원을 검색하고 연결해 주는 레지스트리 시스템을 운영하여 글로벌 데이터 공유 생태계를 주도
- (CHiMaD & MDF) NIST 우수 센터인 CHiMaD(Center for Hierarchical Materials Design)는 'Materials Data Facility(MDF)'를 구축하여 대용량 소재 데이터의 출판, 인텍싱, 클라우드 연동 분석 환경을 제공

□ 특허 데이터의 AI 학습용 전면 개방 및 민관 파트너십

- (개요) 미국 특허상표청(USPTO)은 특허 데이터를 국가 혁신을 위한 핵심 공공재로 정의하고, 민간 기업과 연구소가 이를 AI 학습 및 분석에 자유롭게 활용할 수 있도록 'Open Data Portal'을 통한 전면적인 개방 정책을 추진(출처: USPTO 2022-2026 Strategic Plan)

USPTO launches new Open Data Portal for easy, quick access to data

February 12, 2025

As we modernize our systems and improve customer experience, the U.S. Patent and Trademark Office (USPTO) launched our new [Open Data Portal](#) (ODP) on February 12. In our effort to deliver accessible, open data, this site provides easy access to relevant data sets in a user-friendly site that combines several services into one. Whether you're an experienced researcher, an entrepreneur, or a student, you can quickly access, customize, and extract USPTO data in and from this site.

Support resources are available, including information on [getting started](#) with Application Programming Interfaces (APIs) and more.

ODP will replace both the Patent Examination Data System (PEDs) and the Bulk Data Storage System (BDSS). PEDs will be available for at least 30 days after ODP's launch, and BDSS will be available at least 60 days post launch. More information and specific dates on these retirements will be shared as they become available.

If you'd like to participate in an upcoming feedback session or have questions, please email data@uspto.gov. You can get the latest updates, including details on events related to this topic, by subscribing to receive email updates at our [Subscription Center](#).

출처:United States Patent and Trademark Office (USPTO)

그림 5. USPTO launches new Open Data Portal for easy, quick access to data

□ 핵심 데이터셋 AIPD

- (인공지능 특허 데이터셋(AIPD)) USPTO 수석경제학자실(OCE)은 AI 기술의 발전 양상을 정량적으로 분석하고 연구를 지원하기 위해, 1976년부터 축적된 1,320만 건 이상의 특허 문헌을 머신러닝으로 분석하여 구축한 'AIPD(Artificial Intelligence Patent Dataset)'를 무료로 공개
- (구성) AI 관련 특허 식별 정보, 기술 분류(Vision, NLP, Planning 등 8대 분야), 발명자 정보 등을 포함하며, 연구자들이 별도의 전처리 없이 AI 트렌드 분석에 활용할 수 있도록 최적화
- (활용) 해당 데이터셋은 AI 기술의 확산 속도, 산업별 채택률, 혁신 클러스터 분석 등 다양한 경제·기술 정책 연구의 기초 자료로 활용되고 있음
- 대용량 데이터 제공 및 API 고도화
 - (Bulk Data Service) 특허 전문(Full-text), 서지 정보(Bibliographic), 심사 진행 정보(File Wrapper),

3) MDCS는 사용자 정의 템플릿을 통해 다양한 소재 도메인의 데이터 스키마를 유연하게 생성할 수 있어, 데이터 표준화의 진입 장벽을 낮추는 역할을 수행

특허 양도 정보(Assignment) 등을 매일/매주 단위로 업데이트되는 대용량 파일 형태로 제공하여, Google Patents, Clarivate 등 민간 IP 서비스 기업의 비즈니스 기반을 마련(출처: USPTO Bulk Data Storage System)

- (Patents View) 특허 데이터의 시각화 및 네트워크 분석을 지원하는 플랫폼을 운영하며, 발명자, 기업, 지역 간의 연결 관계를 규명하는 심층 데이터를 제공(출처: PatentsView.org)

○ AI & Emerging Tech Partnership

- (파트너십 결성) AI 기술이 지식재산 시스템에 미치는 영향을 선제적으로 논의하기 위해, 2022년부터 AI 커뮤니티, 학계, 산업계가 참여하는 ‘AI & Emerging Technology Partnership’을 운영(출처: Federal Register Notice 87 FR 38714, '22)
- (정책 피드백) 파트너십을 통해 수립된 의견은 AI 발명자성(Inventorship) 관련 지침 제정, AI 지원 심사 도구 개발 등 USPTO의 주요 정책 수립에 직접적으로 반영(출처: '24년, USPTO Inventorship Guidance for AI-Assisted Inventions)

□ 국가 AI 전략과 R&D 데이터 자산화

- (행정명령 및 전략) 백악관 과학기술정책실(OSTP)은 ‘안전하고 신뢰할 수 있는 AI 개발 행정명령(EO 14110)’ 등을 통해 연방 정부 보유 데이터의 AI 학습용 개방을 가속화(출처: The White House, Executive Order 14110, '23.10)



JOSEPH R. BIDEN, JR.

46th President of the United States: 2021—2025

Executive Order 14110—Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

출처: The American Presidency Project, 2023.

그림 6. Executive Order 14110 on Artificial Intelligence

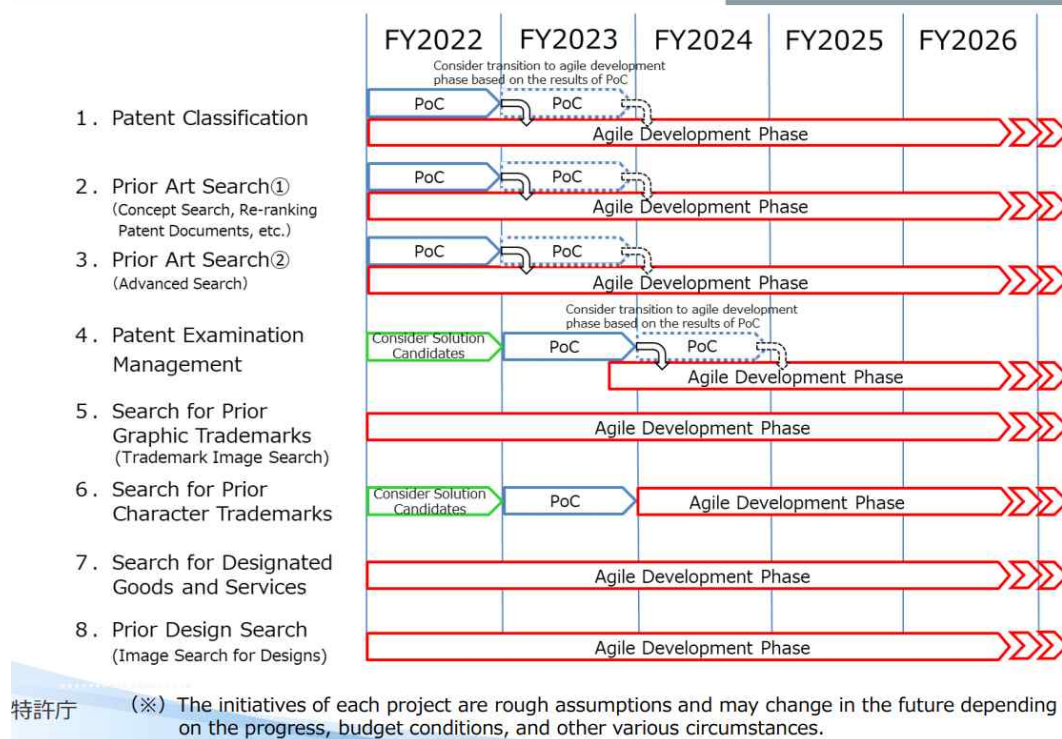
- (AI 학습용 데이터(Assetization)) 연방 기관이 보유한 R&D 성과물(논문, 보고서, 데이터)을 “국가전략 자산”으로 규정하고, 이를 기계 독해가 가능한 형태로 정제하여 공개할 것을 의무화
 - * 이는 중국 등 경쟁국의 데이터 축적에 대응하여 미국의 AI 기술 우위를 유지하기 위한 필수 조치
- (NIST AI 위험 관리 프레임워크(RMF)) AI 데이터의 편향성, 보안, 신뢰성을 관리하기 위한 표준 프레임워크를 개발하여, 데이터 큐레이션 과정에서 준수해야 할 품질 기준을 제시

(일본) 관(官) 주도의 데이터 표준화와 업무 효율화 중심 AI 도입

□ AI 활용 액션 플랜과 ‘정리표준화 데이터’ 정책

- (개요) 일본 특허청(JPO)은 2017년부터 ‘AI 기술 활용 액션 플랜(Action Plan for Utilization of AI Technology)’을 수립하고, 6년 단위의 중장기 로드맵에 따라 특허 행정 전반에 AI 기술을 단계적으로 도입

Action Plan for Utilization of AI technology (FY2022-2026 edition)



출처: JPO, "Action Plan for Utilization of AI Technology", '17

그림 7. Action Plan for Utilization of AI technology (FY2022-2026 edition)

○ AI 도입 로드맵 및 성과

- ('22~'26년 액션플랜) 최근 개정된 계획에서는 특허 분류, 선행기술 검색(Prior Art Search), 상표 이미지 검색(Image Search) 등에 AI 기술을 실증(PoC)하고 실제 업무에 적용하는 'Agile Development' 방식을 채택
- (상표 이미지 검색) 딥러닝 기반의 이미지 인식 기술을 도입하여 도형 상표 검색의 정확도를 높이고 심사 기간을 단축하는 성과를 거둠
- * 심사관들이 AI 검색 도구(Image Search Tool)를 활용하여 선행 상표를 스크리닝함(2024년 기준)
- (특허 심사 지원) AI가 심사관을 대체하는 것이 아니라, 검색 문헌을 관련도 순으로 재배열(Re-ranking)하거나, 키워드 누락을 방지하는 '심사 지원(Assist)' 도구로서의 역할을 명확히 규정하고 개발을 고도화

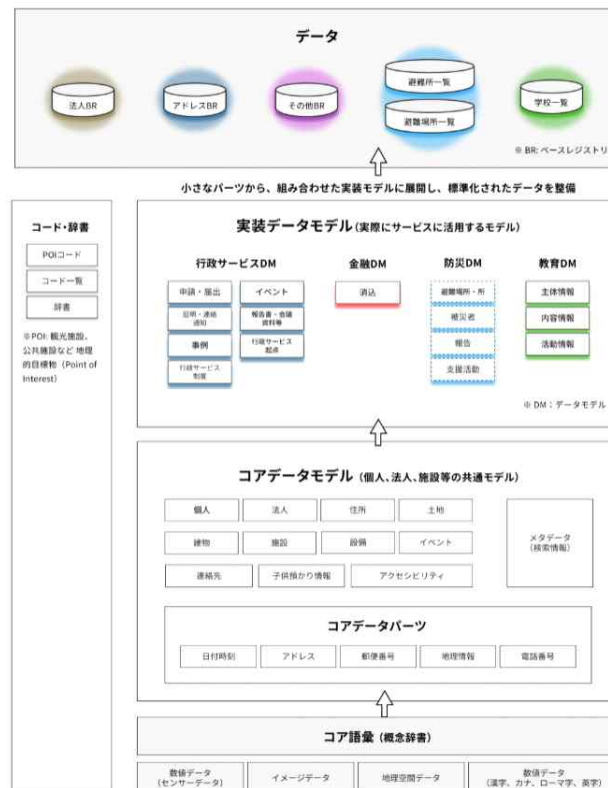
□ 정리 표준화 데이터 기반의 민간 개방 생태계

- (정리 표준화 데이터) 일본은 과거 INPIT(공업소유권정보연수관)를 통해 특허청 보유 정보(서지·경과 정보 등)를 정리표준화 데이터로 정비·제공해 왔으며, 제공 과정에서 데이터의 정리(표준화) 및 형식 변환(SGML → XML 등)이 이루어짐
- * 다만 정리표준화 데이터 제공은 2019년 9월 종료되었고, 이후의 공적 데이터 제공 체계는 특허청(JPO) 중심의 새로운 데이터 제공 체계로 전환됨
- (현행 제공 방식) 일반 대중·실무자를 위한 무료 검색 서비스는 J-PlatPat이 담당하고, 데이터베이스 구축, 서비스 제공 등을 수행하는 민간 사업자 및 연구개발 수요를 위해 특허정보의 일괄 다운로드 서비스 등 별도 제공 채널을 통해 데이터 접근 지원

- (현행 데이터 종류) 정리표준화 데이터 종료 이후, 대표적인 대체제는 특허정보 표준 데이터(서지·경과정보)이며, 이는 특허청이 보유한 특허, 실용신안, 의장, 상표의 서지·경과 정보 등을 업데이트 및 단위로 묶어 제공하는 벌크 데이터로, 민간의 특허정보 DB 구축 등에 활용됨
- * 제공 형식은 TSV 형태로 명시되어 있으며, 이용을 위해서는 다운로드 서비스 등록이 필요하다고 안내됨

□ 데이터 사회 추진과 베이스 레지스트리

- (디지털청의 역할) 정부의 포괄적 데이터 전략 추진을 위해 행정기관 간 정보 연계와 민간의 데이터 활용을 촉진하고 베이스 레지스트리(Base Registry)를 사회 기반 데이터군으로 정의, 해당 데이터의 정비 및 활용 확산 추진
- 디지털청은 베이스 레지스트리를 통해 행정절차에서 원스-온리(Once Only: 한 번 제출한 정보는 반복 제출하지 않도록 하는 원칙) 구현을 지원하고, 데이터 연계를 기반으로 한 민관 협력형 DX 추진을 주요 정책 효과로 제시함
- 베이스 레지스트리는 국가 차원에서 관리하고 활용되는 사회 기반 데이터군으로, 디지털청 자료에서는 문자(표기), 법인, 주소, 공공시설, 지원제도, 이벤트 등을 대표 범주로 예시하고 있음
- 또한 디지털청은 정부 정보시스템 간 상호운용성 제고를 위해 정부 상호 운용성 프레임워크(GIF) 및 관련 API 가이드라인 등을 제시함으로써, 기관 간 데이터 연계 및 민간 활용이 가능하도록 하는 기술·운영 기반(거버넌스 포함)을 마련하는 방향으로 정책을 추진함



출처: 일본 디지털청(Digital Agency), Government Interoperability Framework(GIF)

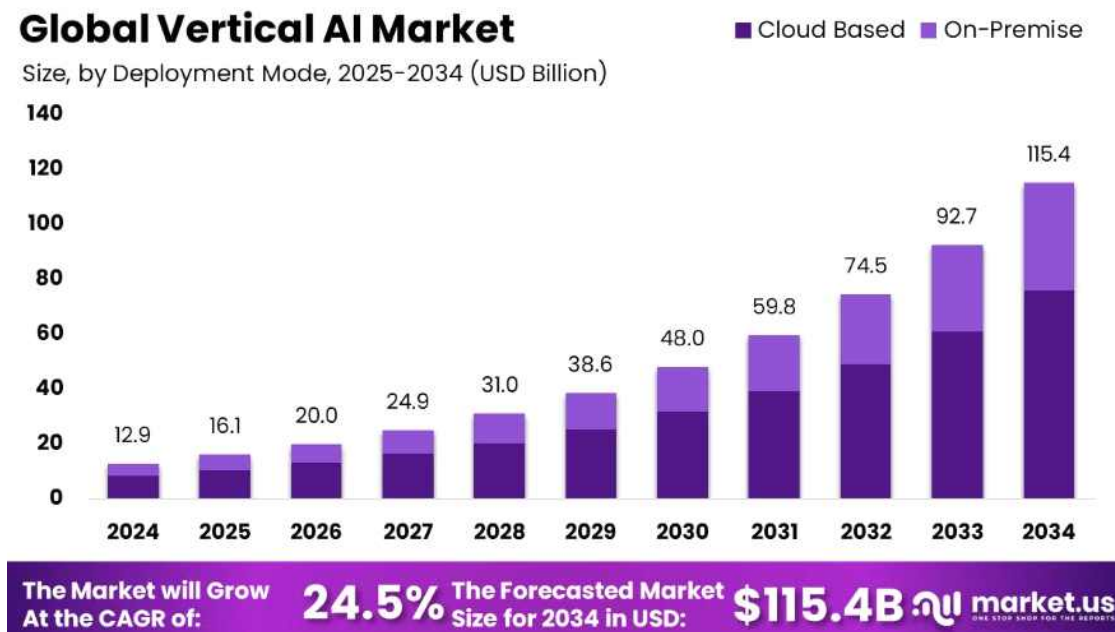
그림 8. 정부상호운용성 프레임워크(GIF) 데이터 모델 구조

제2절 시장 동향

(버티컬 AI) 시장의 급부상과 데이터 중심 생태계

□ Vertical AI 시장 개요 및 전망

- 범용 AI(General AI)의 한계를 넘어서, 특정 산업(Vertical)과 도메인에 특화된 데이터를 학습하여 전문적인 문제를 해결하는 ‘Vertical AI’ 시장이 급성장
- 기업들은 운영 효율성 극대화와 데이터 기반의 정밀한 의사결정을 위해 범용 모델 대신 자사 산업에 최적화된 버티컬 AI 솔루션을 선호하는 추세
- * 글로벌 Vertical AI 시장은 2024년 약 129억 달러에서 연평균 24.5% 성장하여, 2034년에는 약 1,154억 달러 규모에 달할 것으로 전망 (출처: Market.us, 2025)



출처: Market.us, Vertical AI Market Size, Share & Trends Report, 2025.

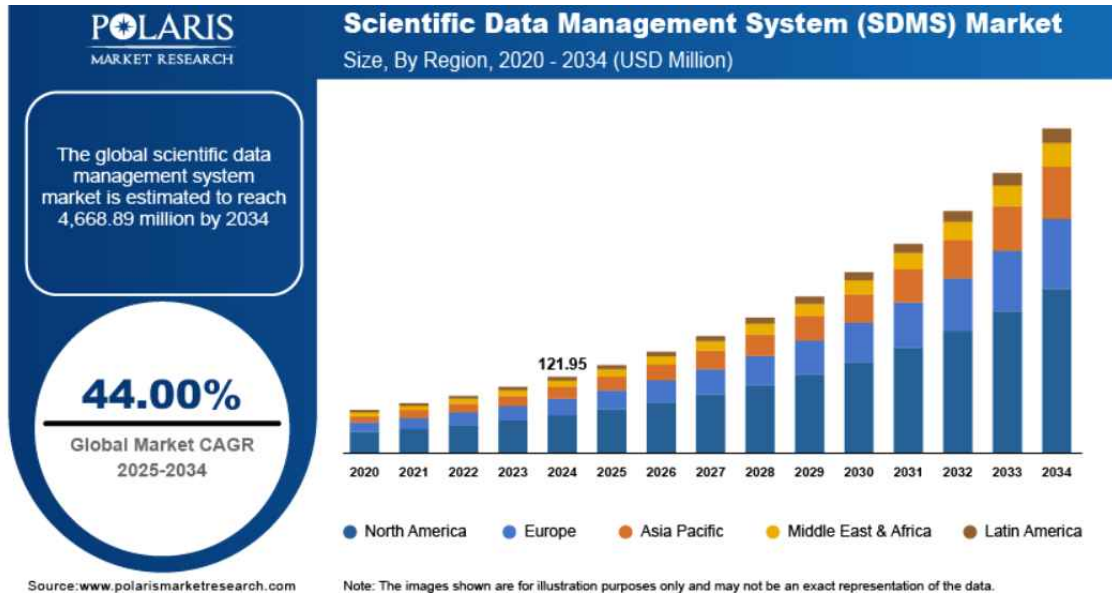
그림 9. Market Size Forecast 2024-2034

□ ‘High-Quality Data’ 확보 경쟁 심화 및 ‘Data-Centric AI’ 부상

- AI 모델의 성능 경쟁(Model-Centric)이 한계에 다다르면서, 고품질 데이터의 확보가 AI의 성패를 가르는 데이터 중심 AI(Data-Centric AI) 시대로 진입
- 웹 문서 등 범용 데이터는 풍부하나, 과학기술 분야의 실험 데이터, 공정 레시피 등 도메인 특화 데이터(Domain Knowledge)는 희소성이 매우 높아 확보 경쟁이 치열함
- 세계적 AI 석학이자 구글 브레인(Google Brain)의 공동 설립자인 앤드류 응(Andrew Ng) 스탠퍼드대 교수는 “AI 시스템의 성능을 높이기 위해서 모델을 고정하고 데이터를 개선해야 한다”며 고품질 데이터 확보의 중요성을 강력히 역설함

□ R&D 데이터 자산화 및 관리(SDMS) 시장의 폭발적 성장

- AI 기반 연구개발(AI for Science)이 보편화됨에 따라, 실험·특허 데이터를 체계적으로 수집·정제하여 AI 학습용 자산으로 전환하는 ‘과학 데이터 관리 시스템(SDMS)’의 중요성이 급부상
- 특히 제약, 바이오, 소재 등 첨단 R&D 분야에서는 비정형 데이터를 기계 독해가 가능한(Machine readable) 형태로 큐레이션 하는 것이 AI 도입의 필수적 전제 조건으로 인식되어 관련 투자가 급증하고 있음
- 글로벌 과학 데이터 관리 시스템(SDMS) 시장은 2024년 약 1.2억 달러에서 연평균 44.0%의 기록적인 성장세를 보이며, 2034년에는 약 46.7억 달러 규모로 확대될 것으로 전망(출처: Polaris Market Research, 2025)



출처: Polaris Market Research, Scientific Data Management System Market Size & Share Report, 2025-2034, 2025.

그림 10. Scientific Data Management System Market Growth

제3절 기술 동향

1 비정형 데이터 처리 기술

OCR·LLM·멀티모달 AI 융합을 통한 고도화된 지능형 문서처리 파이프라인 구현

- 비정형 데이터 처리 기술은 OCR, LLM, 멀티모달 AI를 융합하여 실험 데이터의 의미를 자동 추출하고 지식화하는 전주기 파이프라인으로 정의
- 비정형 실험 데이터 처리 기술은 문서 지능(Document Intelligence) 혹은 지능형 문서 처리(Intelligent Document Processing, IDP)의 하위 분야임
- 광학 문자 인식(OCR), 대규모 언어 모델(LLM), 멀티모달 AI를 결합하여 문서 내 시각적·논리적 정보를 자동 추출하고 구조화하는 기술 체계를 의미함
- * 학술적으로는 이미지 내 텍스트 영역 탐지, 인식, 레이아웃 분석 및 표 이해를 포괄함

- * 산업적으로는 실험 장비 원천 데이터(Raw Data)로부터 핵심 변수(Variables), 단위(Units), 수치 및 상관관계를 추출하여 지식 그래프(Knowledge Graph)나 기계 학습용 데이터 세트로 변환하는 기술로 정의할 수 있으며, 특히 과학기술 데이터는 화학식, 물리 기호, 다차원 도표 등 고도의 전문 지식을 포함하므로 도메인 특화 이해 능력이 필수적임

○ 비정형 데이터 처리는 이미지 전처리부터 추론, 정규화 등의 단계로 진행됨

표 2. 비정형 데이터 처리단계 예시

기술구분	핵심 기능 및 학술적 정의	주요 기술 요소
이미지 전처리 (Preprocessing)	문서 이미지 품질 개선 및 분석 준비	이진화, 노이즈 제거, 기울기 보정
레이아웃 분석 (Layout Analysis)	문서 내 구성 요소(텍스트, 표, 수식 등) 영역 식별 및 분류	객체 탐지, 세그멘테이션
구조 및 내용 인식 (OCR & Structure)	문자 인식 및 표 구조(행, 열)의 논리적 복원	트랜스포머 기반 OCR, TEDS 최적화
멀티모달 추론 (Multimodal Inference)	그림(그래프, 차트)의 시각적 특징과 의미론적 결합 분석	비전-언어 모델(VLM), 시각적 질의응답(Visual QA)
데이터 정규화 및 통합 (Normalization)	추출 데이터의 단위 변환, 오타 교정 및 지식 베이스 매핑	단위 정규화, 에이전틱 RAG

□ (PEST 분석) 사회적 인구 구조 변화와 기술적 진보, 경제적 가치 창출 요구 및 정책적 의지가 결합되어 기술 도입이 가속화됨

○ (사회적 관점Social) 산업계 전반에서 단순 자동화(DX)를 넘어선 고도화된 지능형 의사결정(AX) 체계로의 전환이 요구됨에 따라, 전문 지식이 집약된 비정형 데이터 처리의 사회적 가치가 급증

- (DX에서 AX로 패러다임 전환) 과거의 DX가 정형 데이터를 기반으로 한 업무 절차의 기계적 자동화에 초점이 있다면, 현재의 AX는 비정형 데이터를 통한 지능화된 의사결정과 프로세스 완결을 지향함

- (전문분야의 전문 지식 자산화) 의료, 금융, 제조 등 고도의 전문성이 요구되는 분야에서는 정형 데이터만으로는 포착할 수 없는 핵심 지식이 대부분 문서, 이미지, 오디오 등 비정형 형태로 존재함

- * 특히 의료 기록 내 임상 데이터 분석이나 제조 현장의 비정형 센서 로그 분석 등 전문 영역의 지능화를 달성하기 위해 비정형 데이터 처리 기술이 필수적이라는 사회적 인식이 확산됨

○ (기술적 관점Technological) 멀티 모달리티 및 지능형 분석 인프라 고도화에 따라 비정형 데이터에서 유의미한 패턴을 추출하여 환경 및 산업 관리에 투영하는 기술이 핵심으로 부상

- (멀티모달 및 에이전틱 AI의 생산 환경 정착) 텍스트, 이미지, 오디오를 복합 처리하는 멀티모달 AI가 표준 기술로 자리잡고 있음

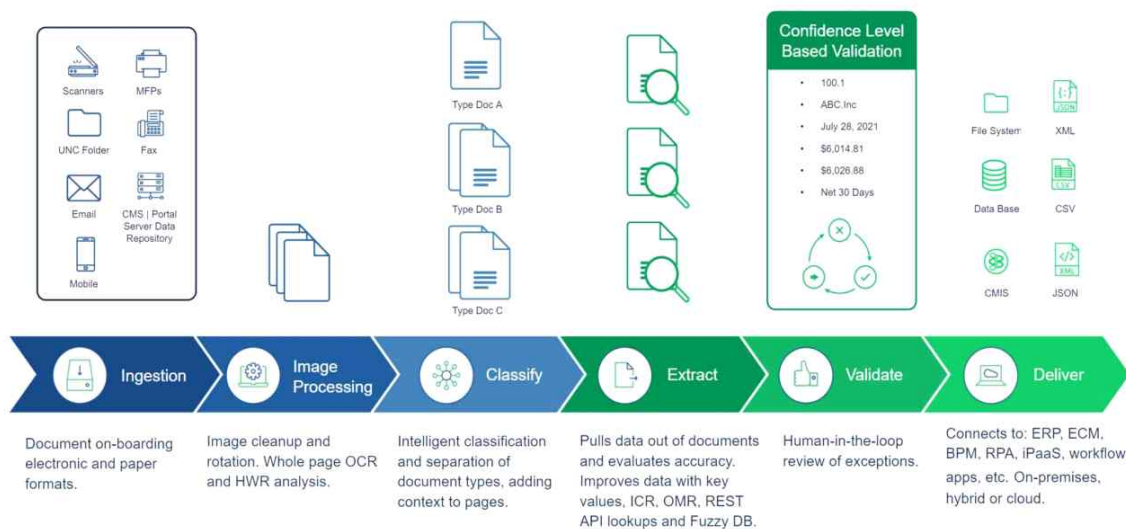
- * 최근 기업용 애플리케이션의 대다수가 자율적 과업 수행이 가능한 AI 에이전트를 탑재할 전망이며, 에이전트 간 표준 통신 프로토콜을 통해 비정형 데이터 기반의 자율 운영 시스템이 구축될 예정임

- (차세대 비정형 데이터 관리 아키텍처 및 인프라 효율화) 데이터 레이크하우스, 벡터 데이터베이스, 지식 그래프 등 비정형 정보의 의미론적 관계를 정의하는 인프라 기술이 중요해지고 있으며, 데이터 폭증에 따른 에너지 문제를 해결하기 위해 저전력 AI 반도체(NPU) 도입 및 친환경 데이터 센터 실증 등 인프라 마련이 진행되고 있음

- **(경제적 관점Economic)** 비정형 데이터 관리 시장의 고성장과 함께 투자의 중심이 ‘버티컬 AI’로 이동하고 있으며, 성과 기반의 비즈니스 모델과 효율적인 비용 관리를 통해 데이터 자산의 경제적 가치가 극대화되고 있음
- **(시장의 폭발적 성장)** 비정형 데이터 관리 시장은 연평균 7.28%의 성장률을 기록하며 2031년 233억 달러 규모를 돌파할 것으로 전망되며, 특히 기업 데이터의 74%가 비정형 형태로 축적됨에 따라, 글로벌 자본은 범용 AI를 넘어 특정 산업 특화모델로 집중되는 추세임
- **(정책적 관점Political)** 비정형 데이터는 국가 안보 및 경제 주권의 중핵으로 인식됨에 따라 범국가적 차원의 법적·제도적 대응이 강화되고 있음
- **(디지털 주권 및 데이터 현지화 규제 강화)** 유럽연합(EU)을 중심으로 ‘디지털 주권’ 확보를 위한 법적 프레임워크가 심화되고 있으며, 국방·의료 등 민감 비정형 데이터의 역내 저장 및 처리를 강제하는 데이터 현지화 정책 확산 중
- **(초거대 AI 및 전문 데이터 육성 전략)** 정부는 품질 데이터 인프라 구축에 주력하고 있으며, AI 학습용 데이터 구축을 위해 전문 분야의 비라벨링 비정형 데이터를 대규모로 확보하여 민간의 AI 도입 장벽을 낮추고 있음
- * 또한 데이터 품질인증 및 가치평가 제도를 시행하여 비정형 데이터의 유통·거래 활성화를 제도적으로 뒷받침하고 있음

□ 비정형 데이터 처리의 상위 개념인 지능형 문서 처리는 「전처리4) → 데이터 추출 → 데이터 검증 → 데이터 적재」를 거쳐 진행

- 사람이 문서 내용을 분석하여 전산 시스템에 내용을 입력하는 방식 등을 AI 솔루션 활용으로 대체하여 컴퓨터 시스템이 완전 자동 처리하게 하는 것이 궁극적인 목표



출처: Kofax

그림 11. 지능형 문서처리 프로세스

- 업무환경에 지능형 문서 처리 도입 시, 수작업 문서 처리 및 인적 오류 발생 가능성을 최소화할 수 있어 기존 시스템 대비 비용 감축 및 정확도 향상 기대
- 문서 당 1분 내로 빠르게 처리할 수 있으며, 인적 오류 최소화 등을 통한 전체적인 문서 처리 오류 감소를 기대할 수 있음

4) 문서 수집, 문서 이미지 처리, 문서 분류 등

- 사람의 검증이 반드시 필요한 규칙 기반 시스템과 달리 신뢰도가 낮다고 판단되거나 중요한 데이터에 한해서만 사람의 검토가 필요한 구조를 갖기 때문에 인건비 절약을 기대할 수 있음

표 3. 문서 처리 방식 비교

구분	수작업 문서처리	규칙 기반 문서처리	지능형 문서처리
처리 시간	문서 당 15분 이상	문서 당 5분 이하	문서 당 1분 이하
정확도	평균 70%	85~90%	98% 이상
사람 개입 수준	전 과정	데이터 검증	일부 데이터 검증
문맥에 따른 데이터해석	가능	불가	가능
복잡한 문서 처리	가능	불가	가능
다양한 문서 형태 처리	가능	불가	가능

- 최근 로보틱 프로세스 자동화(Robotic Process Automation, RPA) 등과 통합하여 전사적인 업무 자동화 솔루션을 제공하는 방향으로 발전하고 있으며, 최신 AI 기술을 적용하여 정확성을 지속적으로 개선하고 있음

□ 지능형 문서처리 프로세스

- **(전처리)** 컴퓨터가 이해하기 쉬운 형태로 문서를 처리하는 과정으로, 크게 문서 수집, 문서 이미지 처리, 문서 분류의 3가지 작업을 수행함
- **(문서 수집)** 다양한 형태의 문서를 스캐너, 팩스, 이메일, 휴대폰, 카메라 등의 여러 채널을 통해 수집하고, 이를 디지털 이미지(PDF, PNG 등) 형태로 변환함
- **(문서 이미지 처리)** 수집 과정을 통해 이미지 형태로 변환된 문서는 노이즈 제거⁵⁾, 이진화⁶⁾, 기울기 보정⁷⁾ 등의 수정 과정을 거쳐 데이터 인식성을 높임
- **(문서 분류)** 쪽 번호, 크기, 모양, 문자의 배치 등 전체적인 형태를 인식하여, 문서의 종류, 순서 등을 구분하고 문서를 분류함
- **(데이터 추출)** 전 처리된 문서 이미지 내용을 광학문자인식⁸⁾, 지능형문자인식⁹⁾ 등의 방법을 사용하여 컴퓨터가 인식할 수 있는 형태로 변환하는 디지털화 작업 수행
- 문서 이미지 내 데이터 추출은 문자가 존재하는 영역을 인식한 후, 해당 영역 내 각각의 문자를 패턴 매칭¹⁰⁾ 혹은 특징 추출¹¹⁾ 기법을 통해 판별하는 방식으로 진행

5) 노이즈 제거는 문서 내용에 해당하지 않는 열록, 줄무늬, 워터마크, 스탬프 및 기타 표시 등을 각종 필터링 기법 등을 활용하여 제거하는 작업

6) 이진화는 어떠한 대상을 일괄적으로 1 혹은 0 같이 둘 중 하나로 분류하는 작업을 의미하는 것으로 지능형 문서 처리에서는 문서를 흑백 형태로 변환

7) 기울기 보정은 기울어진 문서 혹은 문서 스캔 과정 시 기계 오작동 등으로 인해 문서가 기울어진 상태로 인식되었을 때, 올바른 방향으로 보정해주는 작업

8) 광학문자인식(Optical Character Recognition, OCR)은 사람이 쓰거나 인쇄한 문서, 촬영된 사진이나 스캔된 이미지 내의 문서를 인식하여 컴퓨터가 읽고 편집할 수 있는 디지털 텍스트로 변환하는 기술

9) 지능형 문자인식(Intelligent Character Recognition, ICR)은 AI 기술을 활용하여 문자별 고유한 패턴을 파악하고 이를 비교하여 보다 정확하게 문자를 판독하는 방법

10) 미리 준비된 몇 개의 표준 패턴 문자(Template Character)와 입력문자를 각각 일정한 격자 형태 등으로 분해 후 비교하여 표준 패턴 문자 중 가장 유사한 문자로 판별하는 방법

11) 문자를 특정한 형태(원형, 가로·세로 방향에서의 두께 등)로 분해하여 각 성분을 분석 후 가장 유사한 형태의 문자로 판별하는 기법

- 합성곱신경망(Convolutional Neural Network, CNN), 순환 신경망(Recurrent Neural Network, RNN) 등 다양한 AI 기술을 활용하여 문자의 판독 작업을 수행함
- * ICR 기술의 특징 : 비정형적 문자 인식에 있어 OCR 방식 대비 우월한 모습을 보이며, AI 모델 학습 과정이 필요하고, OCR 방식 대비 연산량이 많은 편으로, 동작 시간 및 비용적인 면에서 열위한 모습을 보임

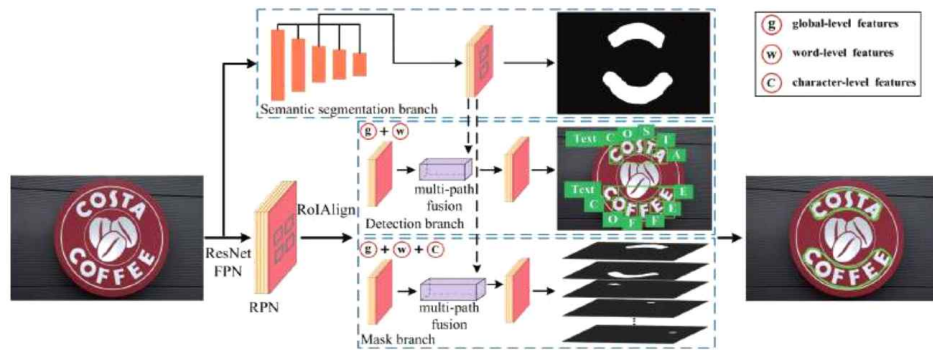


그림 12. 지능형 문자인식 예시

- OCR 및 ICR 기술은 개별문자의 인식에서는 높은 정확도를 보이나, 오기재, 훼손, 변형 등으로 인해 다른 문자로 인식되는 수준일 경우 이를 수정하지 못하는 한계점이 있음
- * 문자인식 오류 발생 여부를 컴퓨터 스스로가 판단하기에 어려움이 있어 데이터 추출 결과물에 대해 사람이 검증하고 확인하는 절차가 반드시 필요함
- 지능형 문서 처리에서는 주로 광학문자인식 혹은 지능형 문자인식에 자연어 처리(Natural Language Processing, NLP)를 조합한 방식을 활용하여 문자를 인식하고 데이터를 추출함
- * 문서별 자주 사용하는 용어, 문서를 작성한 사용자의 의도, 인식된 문자의 전후관계 문맥 등을 고려하여 문자 및 단어를 인식하고, 필요 시 인식과 동시에 교정할 수 있음



그림 13. 자연어 처리 적용 예시

- **(데이터 검증)** 추출된 데이터 중 일부를 사람이 검증하고, 검증 결과를 AI 모델에 학습시켜 품질 개선
 - 중요한 의사결정에 사용되거나, 데이터 추출 과정에서 신뢰도가 낮다고 판단된 데이터에 대해서 사람이 직접 검토하고 수정하는 과정으로 진행
 - 검증결과는 전 과정에 활용된 AI 모델 품질 개선에 사용되며, 검증이 거듭될수록 해당 업무에 대한 지능형 문서처리 정확성 향상으로 이어지게 됨

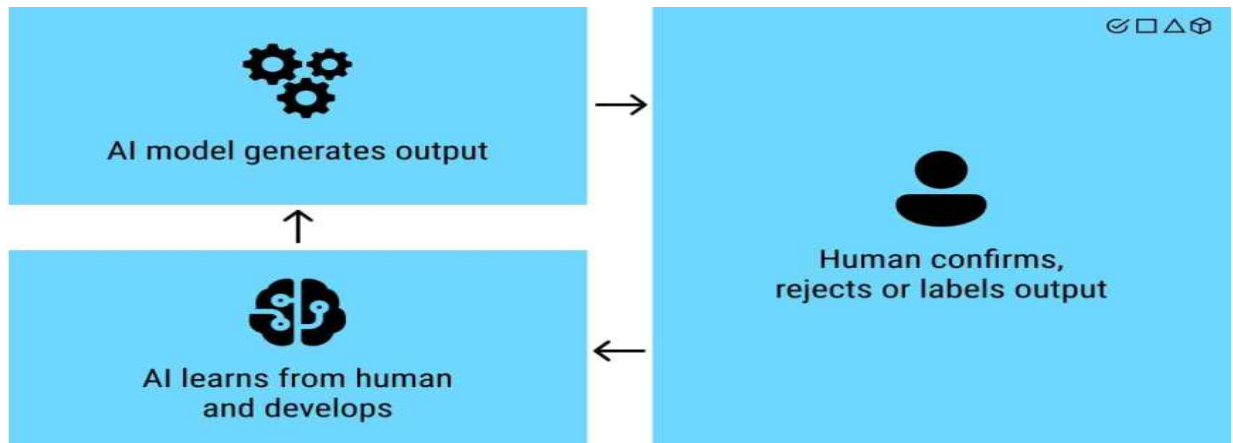


그림 14. 데이터 검증 및 AI모델 학습 흐름도

- **(데이터 적재)** 추출 후 검증 과정까지 완료된 데이터를 각 전산 시스템에 맞게 정형, 반정형 데이터 형태로 변환하여 적재함
 - 적재된 데이터는 각종 업무에 활용될 수 있으며, 데이터 분석, AI 학습 등에 활용되어 전사적인 관점에서 인사이트를 도출하는 것에 이용될 수 있음

□ 업무 자동화 솔루션의 기술 발전 방향

- 전사적인 업무 자동화 솔루션을 제공하는 방향으로 발전하고 있으며, 최신 AI 기술을 적용하여 정확성을 지속적으로 개선하고 있음
 - 타 업무 자동화 솔루션과의 연계, 통합을 통하여 지능형 프로세스 자동화(Intelligent Process Automation, IPA)와 같은 전사적 업무 자동화를 추구하는 방향으로 발전하고 있음
 - GPT-4, Bard, LLaMa와 같은 대규모 언어모델(Large Language Model, LLM¹²⁾)기반 AI를 활용하는 등, 최신 AI 기술을 활용해 전체적인 처리과정에서의 정확도를 개선하고 있음

□ 이미지 분석을 위한 딥러닝 기술의 진화 및 특화 모델 동향

- 최근 이미지 분석 기술은 이미지의 단순 형태 파악을 넘어, 이미지 내 요소 간의 복잡한 구조적 관계를 이해하는 방향으로 진화하고 있음
 - 과거 10여 년간 이미지 분석의 표준이었던 합성곱 신경망(CNN)은 국소적 특징(Local Features) 포착에는 탁월했으나, 도면 전체의 맥락이나 요소 간의 장거리 의존성을 파악하는 데는 한계가 있었음(Krizhevsky et al, 2012)

12) LLM은 방대한 양의 텍스트 데이터를 사전 학습하여 다양한 언어 현상과 지식을 습득한 AI 모델을 말하는 것으로, 문단 수준 이상에서 자연어를 처리할 수 있는 것으로 전망되며, LLM을 적용한 지능형 문서 처리 솔루션은 데이터 추출 과정에서의 정확도가 향상되어 사람의 개입을 더욱 최소화할 수 있을 것으로 기대

- 이런 한계를 극복하기 위해 도입된 비전 트랜스포머(ViT)는 이미지를 패치 단위로 분할하고 셀프 어텐션(Self-Attention) 메커니즘을 적용하여 이미지 전체의 구조적 관계를 학습함(Dosovitskiy et al, 2020)
- * 구조적 데이터 추출(Donut) : OCR 엔진의 의존도를 낮추고 이미지에서 직접 텍스트와 구조를 파악하는 모델. OCR 단계의 오인식 오류가 데이터 추출로 전이되는 것을 방지하여 복잡한 이미지 내 캡션과 수치의 인식 정확도를 개선함(Kim et al, 2022)
- * 특허 전용 멀티모달 분석(PatentLMM): 특허 전용 시각 인코더와 도메인 최적화 언어 모델을 결합한 형태로, 특허 도면 특유의 선(Line)의 의미와 논리적 구조를 일반 모델보다 정밀하게 분석하여 도면 설명을 생성하거나 분석함(Shukla et al, 2025)
- * 관계 및 위상 분석(GNN) : 이미지 내 구성 요소(노드)와 연결 관계(엣지)를 그래프 형태로 학습. 이는 반도체 회로망이나 소재의 원자 구조와 같이 복잡한 계통적 연결성을 이해하는 데 필수적인 기술임(Kipf & Welling, 2017; Li & Chen, 2023)

2 R&D 분야 AI 적용 기술

□ 화학 R&D 지식 자산화를 위한 AI 지능형 추출 기술

- **(MatSKRAFT) 그래프 신경망(GNN) 기반 표 데이터 구조화 프레임워크**, 특허 및 학계 문헌 내 수치 정보가 집중된 표(Table)의 복잡한 구조와 생략된 맥락을 AI가 스스로 파악하여 고차원 정형 데이터로 변환하는 기술
 - **(그래프 구조 변환 및 문맥 이해)** 표의 개별 셀(Cell), 헤더(Header), 캡션(Caption)을 노드(Node)로 설정하고 이들 간의 관계를 엣지(Edge)로 정의하는 그래프 구조를 채택. 이를 통해 표의 기하학적 형태를 넘어 데이터 간 계층적·의미론적 연결성을 파악함
 - **(물리 법칙 기반 제약 조건 인코딩)** 모델 아키텍처에 제약 조건 중심 GNN (Constraint-driven GNN)을 직접 반영하여 추출 수치가 화학적·물리적 개연성을 충족하는지 실시간으로 검증함
 - **(에이전틱 VLM 결합을 통한 정밀도 향상)** 비전-언어 모델(VLM)을 활용하여 이미지 형태의 복잡한 표와 수치 단위를 인식하고, 약 100~200℃와 같은 모호한 범위를 기계가 학습이 가능한 정밀 수치로 정규화함
- **(통합 텍스트 마이닝)** 반지도 학습 기반 개체 인식 및 물성 정규화 기술과 관련하여 ‘Chem Data Extractor’의 정규화 기능과 Snowball 2.0의 고정밀 마이닝 기술을 통합하여, 대규모 비정형 문헌에서 화학 개체와 물성 정보를 누락 없이 수집하고 표준화하는 파이프라인을 구성할 수 있음
 - **(화학 개체명 인식 및 표준 단위 매핑)** 비정형 텍스트 내에서 화학적 개체(Chemical Entities)를 식별하고 관련 물성치를 자동 추출함. 특히 ‘room temperature’, ‘25°C’, ‘298K’ 등 상이한 표현을 내부적으로 표준 SI 단위로 매핑하여 통합 데이터베이스(DB)를 구축함
 - **(반지도 학습 기반 재현율 극대화)** Snowball 2.0 기술을 적용하여 소량의 전문가 주석 데이터만으로도 대규모 특허 군에서 정보를 추출함. 신약 개발 사례에서 기존 시스템 대비 재현율(Recall)을 15~20% 향상시켜 희귀 화합물의 활성 수치(IC₅₀)와 같은 핵심 자산의 포괄성을 확보함
 - **(R&D 공정 최적화 기여)** 온도, 압력, 용매 농도 등 공정 조건의 미세한 차이를 정밀하게 구조화함으로써, 연구자가 과거의 실험 데이터를 기반으로 최적의 반응 조건을 설계할 수 있는 기반을 제공함

□ (인실리코 메디슨) Pharma.AI 플랫폼 및 PACE 엔진 기반 신약 개발 혁신

- (개요) 타겟 발굴부터 임상 단계까지 전 과정을 생성형 AI가 주도하는 ‘End-to-End’ 파이프라인을 구축하고, 특허 분석 엔진 PACE를 통해 비정형 데이터의 지식화를 실현함



출처 : <https://pharma.ai/>

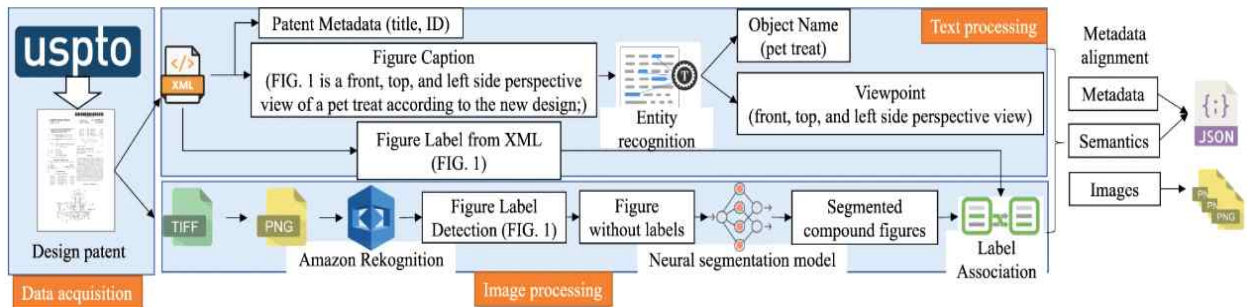
그림 15. 특허 분석 엔진(PACE)

- (특허 분석 엔진 : PACE) 제약·바이오 분야 특허 문서의 심층 분석을 위해 설계된 AI 엔진으로, 단순 텍스트 검색을 넘어 비정형 문서의 구조화에 특화
 - 시각 지능 기반 데이터 추출 : 비전 모델과 비주얼 언어 모델(VLM)을 활용하여 PDF 내 이미지 형태의 화학 구조식 및 SAR(구조-활성 상관관계) 테이블을 자동 인식하고 디지털 데이터로 변환함
 - 특허 공간(Patent Space) 분석 : 추출 데이터를 기반으로 구조-활성 관계를 분석하고, 기존 특허들 사이의 공백 영역 등을 파악하여 신규 물질 설계의 방향성을 제시함
- (통합 플랫폼 성과) 핵심 엔진인 PandaOmics(타겟 발굴), Chemistry42(분자 설계), inClinico(임상 예측)를 유기적으로 연결하여 신약 개발 소요 기간을 획기적으로 단축함
 - 타겟 발굴 및 후보물질 도출 : 폐섬유증 치료 타겟인 TNIK 단백질 발굴 및 최적화된 분자 구조 생성을 포함, 약 50일 만에 후보물질(ISM001-055) 도출 및 합성에 성공함
 - End-to-End 공정 가속화 : 타겟 발굴부터 임상 1상 완료까지 약 30개월 만에 완수(전통적 방식 대비 수년의 기간 단축)하였으며, 약 15만 달러의 저비용으로 후보물질 도출 성과를 거둠
- (주요 마일스톤: Rentosertib 개발 사례) 생성형 AI가 설계한 약물(ISM001-055)이 타겟 발굴부터 임상 2상 성공까지 주도한 최초의 사례로 기록됨

- 글로벌 임상 검증: 뉴질랜드와 중국에서 진행된 임상 1상을 통해 안전성 및 약동학 프로파일의 양호함을 입증하였으며, AI 주도 신약 개발의 실효성을 전 세계적으로 증명함
- * 출처 : 바이오인(BioIN), "인실리코 메디슨, End-to-End 생성형 AI가 설계한 약물로 임상 2상까지 최초 성공(2026)" 내용 재구성

□ (DeepPatent2) 대규모 특허 도면 이해 및 지능형 시각 분석 프레임워크

- (개요) 색상과 질감이 부재한 스케치 형태의 특허 도면을 정밀 분석하기 위해 설계된 대규모 벤치마킹 코퍼사이자 분석 파이프라인으로, 복잡한 기계 도면의 시각적 상세도를 지능적으로 구조화하는 기술임



출처 : DeepPatent2: A Large-Scale Benchmarking Corpus for Technical Drawing Understanding

그림 16. The architecture of the pipeline to create DeepPatent2

- (지능형 데이터 구축 및 텍스트/이미지 정합 파이프라인) 특허 도면의 시각적 정보와 XML 명세서의 텍스트 정보를 고도로 결합하여 메타 데이터화함
 - 의미론적 정보 추출(NLP): BiLSTM-CRF와 DistilBERT를 결합한 NER 기법을 통해 도면 설명문에서 객체 이름과 시점(Viewpoint) 정보를 추출하며, F1 스코어 0.960의 높은 정확도를 달성함
 - 고정밀 이미지 분할(CV): 복합 도면(Fig 1, 1a 등) 내 텍스트 레이블을 마스킹한 후, MedT(Medical Transformer) 모델을 적용하여 순수 기하학적 구조 기반으로 도면을 약 99%의 정확도로 개별 분리함
 - 데이터 다양성 확보: 2007년 대비 2020년 개별 도면 수가 약 80% 정도가 증가하는 추세를 반영하여, 인공지능이 현대 산업 디자인의 복잡성을 학습할 수 있는 견고한 데이터 기반을 제공함
- (R&D 시각 분석 및 검색 엔진 고도화 기술) 단순 픽셀 비교 방식에서 벗어나 도면의 개념적 유사성 (Conceptual Similarity)을 파악하는 차세대 검색 기술을 구현함
 - 계층적 다중 긍정 대조 학습 : 로카르노 국제 분류(LIC) 체계를 학습 프로세스에 통합하여, 동일 특허뿐만 아니라 의미론적으로 유사한 하위·상위 분류군 도면들을 가깝게 배치하도록 모델을 최적화함
 - 검색 성능의 비약적 향상: 계층적 접근법과 도면 특화 손실 함수를 통해 기존 모델 대비 평균 정밀도 (mAP)를 33.5% 이상 개선하여 최종 93.5%의 mAP를 기록
 - 강건성 모델 학습: 2.7만 개 이상의 시점 정보를 활용하여 객체가 회전하거나 분해된 상태에서도 동일 부품임을 인식할 수 있는 높은 공간적 이해도를 확보

□ (시트린-파나소닉 사례) 머신러닝 기반 소재 인포매틱스 및 IP 고도화 사례 분석

- (개요) 시트린 인포매틱스(Citrine Informatics)와 파나소닉(Panasonic)은 머신러닝 플랫폼을 활용하여 신소재 발견을 가속화하고, 고성능 화합물에 대한 지식재산권(IP)을 신속히 확보하는 것을 핵심 가치로 삼아 유기 반도체 소재 개발 협업을 추진함

- 연구진의 직관 및 물리 법칙과 인공지능 방법론을 결합하여 100만 개 이상의 후보군 중 유망 타겟을 정밀하게 선별함. 그 결과 기존 대비 전하 이동도가 25% 향상된 신규 분자를 발굴하고 총 4건의 특허 출원을 진행하는 독보적인 R&D 성과를 달성함
- **(추진배경 및 핵심과제)** 차세대 IoT 기기 구현을 위해 유연성, 경량성 및 경제성을 갖춘 반도체 소재 확보가 필수적인 상황에서 파나소닉은 유기 용매에 용해되어 저렴한 스핀 코팅 공정이 가능한 헤테로아센 계열 소재를 검토 중이나, 소재 성능 향상과 공정 비용 절감을 동시에 달성해야 하는 기술적 과제에 직면함
- 다만 유망 후보 화합물의 종류 수가 수백만 개 이상으로 방대하여 기존의 실험적 방식이나 물리 기반 시뮬레이션만으로는 진수 조사가 사실상 불가능한 실정이었기에, 이러한 한계를 극복하기 위해 파나소닉과 시트린 인포매틱스는 각 사의 전문 지식을 결합하여 방대한 탐색 공간 내에서 최적의 후보 물질을 효율적으로 찾아내기 위한 협력을 추진함

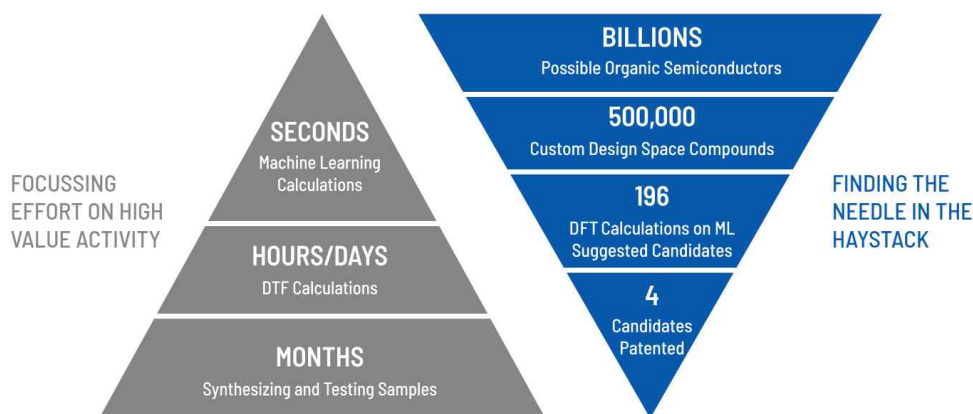


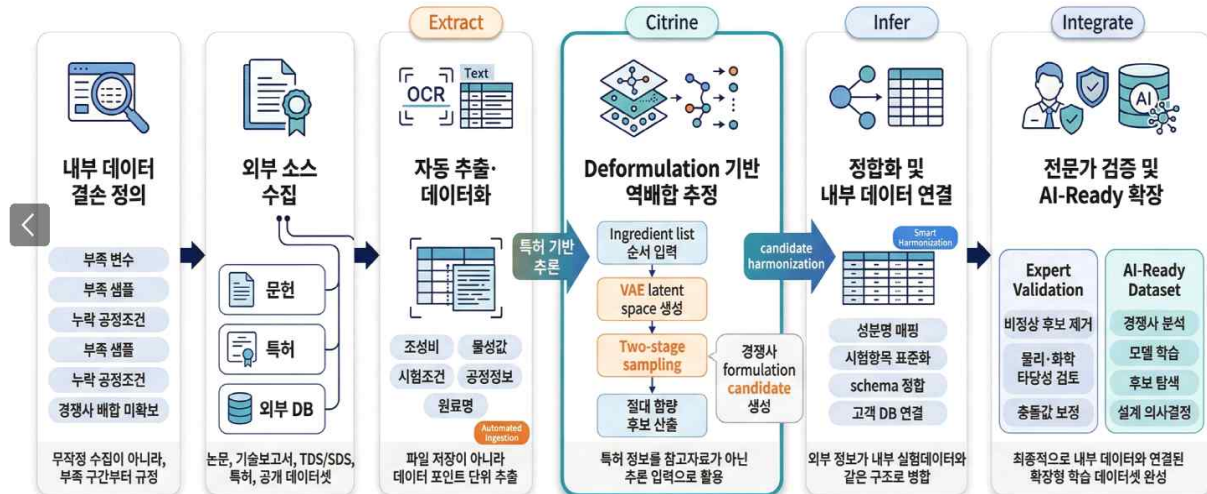
그림 17. 시트린-파나소닉 협업 성과

- **(AI 기반 R&D 전략)** AI 플랫폼 및 순차적 학습 방법론 기반 신소재 R&D 추진
- (인간의 직관과 AI의 결합) 연구진의 도메인 지식과 AI의 화학 구조 해석 역량을 결합하여 50만 개의 실현 가능한 후보군을 정의하고 유의미한 데이터로 전환함
- (하이브리드 워크플로 구축) 머신러닝의 초고속 선별과 물리 기반 고정밀 시뮬레이션을 연계하여 방대한 후보군에 대한 검증 효율을 극대화함
- (순차적 학습 방법론 적용) 실험 설계 단계부터 AI 기반으로 순차적 학습을 도입하여 성공 가능성이 가장 높은 지점을 집중적으로 타격하는 최적화 방식을 채택함

표 4. AI 기반 R&D 전략

구분	기존 방식(Physical R&D)	AI 기반 방식(Informatics)
탐색범위	• 실험 및 시뮬레이션 수행 역량 한계로 인한 제한적 범위탐색	• 수백만 개 이상의 후보군에 대한 광범위한 초고속 스크리닝 수행
소요시간	• 물리적 시료 합성 및 테스트 수행 시 수개월의 기간 소요	• 머신러닝(초단위) 및 DFT 시뮬레이션(수일 단위) 활용으로 연구 기간 획기적 단축
분석도구	• 연구자 직관에 의존한 시행착오 중심의 실험 방식 채택	• 머신러닝 기반 유망타겟의 우선순위 설정 및 실험 설계 최적화

문헌·특허·외부 DB의 정보를 구조화하고, 필요시 deformation으로 경쟁 formulation 후보까지 확장



핵심은 외부 자료를 검색하는 것이 아니라, 내부 데이터의 결론을 메우고 경쟁 formulation 후보까지 AI 학습 가능한 구조로 편입하는 것

문헌·특허·DB 수집 → 데이터 포인트 추출 → deformation 추론 → 정합화 → 전문가 검증

그림 18. 시트린 사의 AI 학습용 특허 데이터셋 구축 프로세스

○ (성과도출) AI 플랫폼 기반 R&D 생산성 혁신 및 지식 자산화 성과 분석

- (R&D 생산성 혁신) 100만 개 이상의 후보 물질 중 단 196건의 DFT 계산만으로 최적 결과를 도출하여 비용 및 시간을 획기적으로 절감함
- (고성능 소재 개발 및 IP 확보) 기존 대비 전하 이동도가 25% 향상된 신규 분자를 설계하고 핵심 화합물 4종의 특허 출원을 완료하여 시장 지배력을 강화함
- (연구 지식의 자산화) 분자 토폴로지 등 과학적 통찰을 머신러닝 모델에 반영하여 향후 프로젝트에 재사용 가능한 디지털 자산으로 전환함

Panasonic-Citrine 유기 반도체 탐색 사례

거대한 분자 탐색공간에서 AI와 계산화학을 결합해 고성능 후보를 빠르게 선별한 구조

1. 산업 배경 및 탐색 문제

개발 목표

- 경량
 - 유연성
 - 저비용 공정
 - 고성능
- IoT 전자소자용 유기 반도체 소재 확보

주목한 분자군

heteroacene 계열
액상 공정 가능
spin-coating 친화
플렉서를 전자소자 적합

핵심 병목

가능 후보: 수십억 개
검토 가능 후보: 수십만 개
계산 가능 후보: 수백 개
특허·실험 가치 후보: 수개

거대한 화학 설계 공간에서 바늘 찾기 문제

2. Citrine의 탐색 구조

설계 공간 축소

연구자 직관 / 도메인지식 → Custom Design Space → 화학식·분자구조 해석

무한 후보가 아니라 화학적으로 타당한 탐색공간 정의

Sequential Learning Loop

모델 업데이트 → 후보 추천 → Citrine AI → DFT / MD 계산 → 결과 수집

계산 결과가 쌓일수록 다음 후보 추천 정확도 향상

시간 계층: Seconds Machine Learning, Hours-Days DFT Calculations, Months Synthesis & Testing

가장 빠른 계층에서 후보를 먼저 제거

AI의 실제 역할

고성능 후보 우선 추천, 불확실성 고려, 계산 자원 집중, 탐색 효율 극대화

Design Space Reduction

Sequential Learning

IP Creation

3. 정량 성과 및 의미

KPI

Hole Mobility: +25%, DFT: 196건, >1,000,000 후보 중 선별

전수조사 없이 유의미한 성능 개선 달성

사업 성과

특허 출원: 후보 4개, IP 자산화

축적된 과학적 인사이트

구조-성능 관계 학습, fused thiophene ring 영향 이해, 향후 프로젝트 재사용 가능, 단발성 추천이 아닌 지식 자산 누적

핵심은 AI가 연구자를 대체한 것이 아니라, 연구자 지식을 바탕으로 탐색공간을 줄이고 계산 우선순위를 최적화해 성능·속도·IP를 동시에 확보했다는 점

거대한 화학공간 → 설계공간 정의 → AI-DFT 반복학습 → 고성능 후보 및 특허 확보

그림 19. 시트린-파나소닉의 AI 학습용 특허 데이터셋 구축을 통한 유기 반도체 후보 탐색 사례

- **(전략적 시사점)** 수십억 개의 화학적 조합 중 최적 물질을 선별하는 방대한 탐색 공간 내 유망 물질 도출 과정에서 AI는 단순한 분석 도구를 넘어 연구 성공률을 비약적으로 제고하는 핵심 엔진 역할을 수행하며 R&D 패러다임을 전환함
- 특히 AI 플랫폼을 통해 도출된 고성능 유망 물질을 즉각적으로 특허화함으로써 강력한 기술 진입장벽을 선제적으로 구축하고, 연구 성과의 상업적 가치를 극대화하여 글로벌 시장 지배력을 확보함

□ **(팔란티어 사례) 데이터 큐레이션과 AI 솔루션 컨설팅을 결합한 하이브리드 서비스 모델 구축**

- **(서비스 모델)** 팔란티어(Palantir Technologies)는 SaaS(서비스형 소프트웨어)에 전문가의 인적 서비스(컨설팅, 운영 대행 등)를 결합하여 제공하는 형태인 SwaS 비즈니스 모델을 운영하고 있으며, 현장 배치 엔지니어(FDE) 모델과 5일 부트캠프 방식을 통해 고객 데이터 기반의 즉각적 가치 입증에 성공하며, 엔터프라이즈 AI 시장을 선도 중
- **(현장 배치 엔지니어(FDE, Forward-Deployed Engineer))** 고객사에 상주하며 실제 프로덕션 환경에서 자사 플랫폼(Foundry, AIP)을 기반으로 커스텀 데이터 파이프라인 구축·배포
 - 이를 통해 고객별 커스터마이징이 플랫폼 제품화로 축적되어, 후속 고객의 도입 비용·기간을 대폭 절감하고 있으며, 단순 컨설팅과의 핵심 차이점은 고객 현장에서 발견한 반복적 문제 패턴을 자사 코어 플랫폼의 표준 기능으로 병합(Upstream)하여, 맞춤형 서비스 경험이 제품 경쟁력으로 전환되는 선순환 구조 형성
- **(AIP 부트캠프(5일))** 전통적 B2B 소프트웨어 영업 주기(12~18개월)를 단 5일로 압축하는 고객 유치 전략으로, 사전 가공 데모가 아닌 고객사의 실제 데이터와 라이브 환경을 기반으로 진행
 - 5일 차에 프로덕션 환경 적용 가능 수준의 솔루션을 도출하여 경영진에게 투자 대비 효과(ROI)를 사전 증명
 - 부트캠프 참여 고객의 유료 계약 전환율 75% 이상, 평균 계약 규모 약 13억 원 이상

표 5. 5일 부트캠프 운영 개요

일차	핵심 활동	주요 산출물
Day 1	• 비즈니스 문제 정의, 데이터 인벤토리, KPI·성공지표 확정	• 업무 흐름도, 성공지표 문서
Day 2	• 핵심 데이터 수집·정제, 품질 리포트 생성, 베이스라인 모델 구현	• 데이터 딕셔너리, 품질 리포트
Day 3	• AI 모델 MVP 구현, 업무 시스템 최소 연동, 성능 평가 수행	• 모델 카드, 평가 리포트, 추론 API
Day 4	• 현업 UAT 검증, 리스크 테스트, 운영 시나리오 설계	• UAT 기록, ROI 1차 분석
Day 5	• 임원 대상 운영 워크플로 시연, 확장 로드맵·SOW·가격 옵션 제시	• 제안서, SOW 초안, 가격 옵션

- (수익구조) 팔란티어의 비즈니스 모델은 서비스로 고객을 유치하고, 플랫폼으로 수익을 창출하는 구조 (전통 IT 컨설팅(프로젝트 종료 시 수익 중단)과 근본적 차별화)

표 6. 3단계 수익 구조

단계	핵심 활동	전략적 목표
1단계 획득(Acquire)	• 부트캠프·파일럿을 저비용으로 제공, FDE 전면 배치하여 고객 실데이터 기반 워크플로를 단기간 내 구현	• 진입 장벽 제거, 코어 플랫폼을 고객 데이터 환경에 이식하여 대체 불가능한 기반 형성
2단계 확장(Expand)	• 인접 부서·유즈케이스·데이터 범위 확장, FDE가 새로운 유즈케이스를 발굴하여 계약 규모 지속 상승	• 플랫폼이 고객 운영에 직조(Woven)되어 전환 비용 상승, 순 달러 유지율(NRR) 134% 달성
3단계 규모화(Scale)	• 반복 패턴을 표준 모듈·산업 패키지로 제품화, FDE 인력 철수 후 소프트웨어 라이선스 수익 극대화	• 매출 총 이익률 80% 이상 달성, 한계 비용 0 수렴, 고객 자생적 운영 단계 진입

- (전략적 시사점) 글로벌 엔터프라이즈 AI 시장은 단순 소프트웨어 판매가 아닌 ‘데이터와 컨설팅(AI 솔루션)을 동시에 제공하는 SwaS 하이브리드 모델’이 표준으로 정착되고 있음

□ (엘스비어 리액시스 활용 론자(Lonza) 사례) IP 데이터 큐레이션 및 AI 예측 모델 기반의 신약 제조 공정 최적화

- (솔루션 모델) 글로벌 정보 분석 기업 엘스비어(Elsevier)는 방대한 IP 및 학술 데이터 자산을 기계 가독성 (Machine-Readable) 형태로 큐레이션하고, 이를 AI 역합성 예측 모델인 ‘리액시스(Reaxys)’와 결합하여 CDMO 기업인 론자(Lonza)의 신약 제조 공정 혁신을 주도
- (데이터 자산화 및 기술 실증) 105개국 특허청 데이터 및 1.8만 종 저널에서 추출한 3억 5천만 개의 물질 정보와 5억 건의 물리화학 데이터를 표준화·구조화하여 데이터 세트를 구축
 - (AI 고도화 및 실증) 7,300만 건의 반응 데이터에 딥 신경망(DNN) 및 몬테카를로 트리 탐색 (MCTS) 알고리즘을 적용한 지능형 솔루션을 론자의 실제 API(활성원료의약품) 제조 현장에 투입하여, 합성 경로를 최대 11단계 축소하고 평균 17%의 공정 단축을 달성함으로써 신약 출시 가속화 및 원가 절감의 선순환 체계 마련

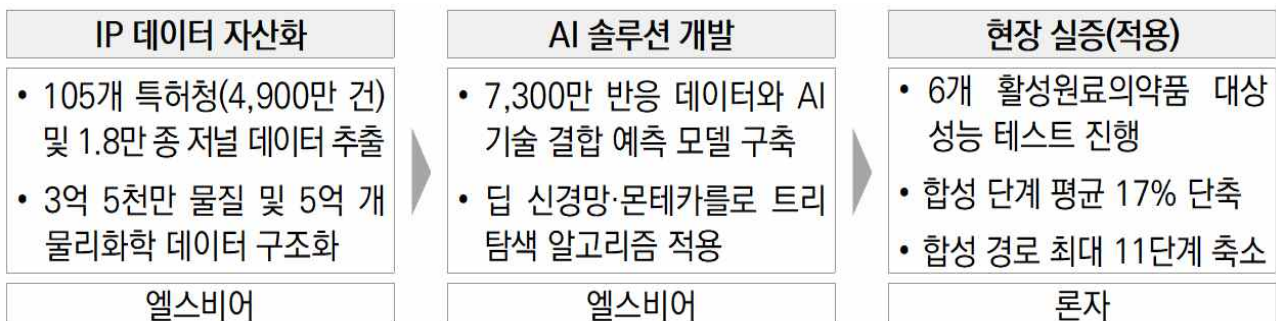


그림 20. 엘스비어 리액시스 활용 사례

제3장 R&D 추진전략

제1절 추진 방향 도출

1 추진방향 설정

□ 사업 기본방향 설정 프로세스

- 지식재산 데이터 큐레이션 기반 고효율 연구개발 사업의 추진 방향을 설정하기 위해 대내외 환경분석 및 기술개발 가능성 등을 종합적으로 분석

표 7. 환경 분석 및 기술개발 가능성 분석

환경 분석 및 이슈진단	현장 수요 및 전문가 검증	전략 수립 및 상세 기획
• 대내외 환경 분석 - 주요국 정책 동향, 기술개발 동향 등 환경 분석 • R&D 가능성 평가 - 국내외 기술현황 조사를 통해 정부 투자의 정책적·경제적 타당성 검증	• 기술 수요조사 - 인터뷰·설문 등 현장 밀착형 과제 요건 및 기획 적합성 도출 • 전문가 검증 - 산·학·연 전문가로 구성된 기획 위원회 운영을 통해 수요 과제 확정	• 추진 전략 설계 - 환경분석 및 수요조사 결과 바탕으로 지원 전략 수립 • 실행 로드맵 수립 - 연차별 데이터 스펙트럼 확장 계획 및 지능형 R&D 로드맵 확정

- **(환경분석 및 이슈진단)** 글로벌 기술 경쟁 심화와 AI 학습용 데이터 고갈 위기 등 대내외 환경변화 분석을 통해 특히 내 비정형 데이터의 활용 저해 요인을 도출하고, 국가 전략기술 확보를 위한 지능형 IP-AI R&D의 추진 당위성을 확보함
- **(기획 적합성 진단)** 사업 추진의 합리성과 객관성을 확보하기 위해 특히 중심의 R&D 성공 가능성을 검증하고, 이후 특히 데이터 및 AI 전문가를 대상으로 체계적 수요 조사를 수행하여 사업의 현장 적합성을 극대화함
- **(추진전략 도출)** 환경분석 및 국내 R&D 역량 분석 결과를 바탕으로 국내 강점(고품질 특허 보유)과 약점(비정형 데이터 처리 미비)을 분석하여, 1단계(PoC) 및 2단계(Scale-up)로 이어지는 단계별 추진 전략을 도출함
- **(전문가 검증)** Top-down(국가 전략기술 연계) 및 Bottom-up(현장 연구자 수요) 방식을 병행하고, 지식재산처 기획위원회와의 점검 및 수요조사 결과 검증을 통해 중점 지원 분야를 최종 선정함
- **(로드맵 상세화)** 환경분석 결과와 지식재산 데이터 큐레이션 기술의 성숙도를 반영하여, 연차별 데이터 스펙트럼 확장 계획 및 지능형 R&D 서비스 로드맵을 포함한 세부 사업 내용을 상세 기획함

2 SWOT 분석

- **(Strengths)** 세계 최상위권의 지식재산 행정 역량과 국가전략 데이터 자산, 고도화된 AI 기술 인프라의 유기적 결합

- 글로벌 최고 수준의 지식재산(IP) 인프라 및 선진적 사법·행정 시스템 보유
 - 2025년 국제지식재산지수(GIPC) 종합 세계 10위 달성 및 특허권 분야 세계 2위 기록 등 글로벌 최상위권의 제도적 환경 입증
 - 2024년 대비 2025년 순위가 6단계 수직 상승(16위→10위) 하는 등 국가 차원의 지식재산 보호 및 활용 환경이 비약적으로 개선됨
- 기술 패권 경쟁의 핵심 무기인 5억 건 이상의 독점적 특허 빅데이터 자산 확보
 - 전 세계 신규 기술 정보의 약 80%가 담긴 특허 문헌은 기술 수명주기와 권리 정보가 결합된 유일한 시계열 국가전략 데이터임
 - 기술적 난제 해결과 법적 보호 전략을 동시에 수립할 수 있는 5억 건 이상의 독점적 기술 정보는 국가 지식재산 총역량을 강화하는 핵심 지표임
 - 원천·경쟁사 특허의 권리 소멸 시점 분석 및 중복 투자 방지 등 R&D 전주기에 걸쳐 활용이 가능한 고가치 정보 자산의 집중도 우수
- 데이터 주도형 혁신을 뒷받침하는 명확한 법적 근거 및 범부처 정책 추진력 확보
 - 「산업재산 정보의 관리 및 활용 촉진에 관한 법률」 제16조 및 제23조를 통해 비정형 데이터를 포함한 정보의 R&D 활용 및 상용화 지원 근거 명시
 - 「지식재산기본법」 및 「발명진흥법」에 의거, 지식재산 서비스업 육성과 국가 R&D 성과 연계 시책을 추진할 수 있는 제도적 기틀 마련
 - 정부의 강력한 기술보호 의지와 「AI+R&DI 추진전략」 등 범부처 합동 대응 거버넌스를 통해 사업의 실행력과 정책적 정당성 확보
- OCR·LLM·멀티모달 AI 등 세계 수준의 지능형 문서 처리 기술 인프라 축적
 - 5G, AI, 블록체인 등 안정적 ICT 인프라를 바탕으로 지능형 R&D 솔루션으로 빠른 기술 전환이 가능한 여건 보유
 - OCR, LLM, 멀티모달 AI를 융합하여 특허 내 비정형 실험 데이터의 의미를 자동 추출하고 지식화 하는 전주기 지능형 문서 처리(IDP) 파이프라인 구축
 - 이미지 전처리, 레이아웃 분석, 구조 인식 및 멀티모달 추론 등 데이터 정규화와 통합을 위한 세부 핵심 기술 요소의 완성도 확보

□ (Weaknesses) 특허 데이터의 높은 비정형성 장벽 및 AI 학습용 고품질 자산 미비

- 특허 고유의 복합적 언어(각 국가 언어로 출원) 장벽 및 비정형 구조로 인한 데이터 활용의 구조적 한계
 - 기술 내용을 법률적 용어로 방어하는 특수한 문체로 인해, 일반적인 자연어 처리(NLP) 기술로는 핵심 정보(목적·구성·효과)를 정확히 추출하고 해석하는 데 근본적 한계가 존재함
 - 연구개발의 필수 정보인 실험 수치, 화학 구조식, 도면 등이 비정형 이미지나 텍스트로 복잡하게 산재해 있어 AI가 직접 학습하기 어려운 데이터 구조적 난제가 상존함
 - 특히 화학·소재 분야의 핵심 데이터 대부분이 표나 이미지 등 비정형 형태로 잠겨 있어, 실제 연구 현장에서 활용되지 못하고 ‘비정형 데이터의 활용 사각’ 상태에 직면해 있음

- 기존 지식재산 인프라의 단순 검색 편중 및 AI 추론용 데이터셋 대응력 미흡
 - 키프리스(KIPRIS) 등 기존 특허 데이터베이스는 키워드 매칭 중심의 단순 검색 기능에 최적화되어 있어, AI의 인과관계 학습 및 추론을 위한 고도화된 학습 데이터셋 제공에는 한계가 노출됨
 - 정부 차원의 국가전략 기술 육성 방안 마련에도 불구하고, 실제 지능형 R&D 현장에서 즉시 활용 가능한 수준의 IP 데이터 자산화 인프라 및 전용 추론 플랫폼은 여전히 미비한 실정임
- 고품질 학습 데이터 구축을 위한 민간의 기술적·경제적 수용력 한계
 - 개별 기업이 독자적으로 감당하기 어려운 고품질 학습용 정답 데이터 부족 현상이 심화되고 있으며, 국가 차원의 전문적인 큐레이션 및 표준 데이터셋 지원이 절실함
 - 민간의 독자적인 데이터 해독 및 자산화 시도는 막대한 구축 비용과 고난도 기술력이 요구되어 정체되어 있으며, 이는 중소·벤처기업의 AI 도입 및 R&D 혁신에 높은 진입 장벽으로 작용함
- 연구 현장의 파편화된 데이터 관리 및 체계적인 공유·표준화 메커니즘 부재
 - 개별 연구실이나 기관 단위로 산재된 비정형 실험값(데이터)은 규격화 및 공유가 어려워 사장되는 경우가 많으며, 이를 공공·민간 특허 빅데이터와 유기적으로 융합 분석할 수 있는 인프라가 부족함
 - 연구자의 개인적 직관과 반복 실험(Trial & Error)에 의존하는 기존 R&D 방식은 기술 난도 상승에 따라 고비용·저효율의 한계에 봉착해 있음
- AI 기반 문서 처리 기술의 태생적 한계 및 전문가 검증 의존도 상존
 - OCR 및 ICR 기술은 개별 문자의 높은 인식률에도 불구하고 오기재, 훼손, 변형 등으로 인한 인식 오류를 스스로 판단하거나 교정하지 못하는 기술적 한계가 여전히 존재함
 - 데이터 추출 결과물의 신뢰성을 확보하기 위해서는 도메인 전문가의 수동 검증 및 확인 절차가 반드시 수반되어야 하므로, 완전 자동화 달성까지는 일정 수준의 시간과 인건비 투입이 불가피함

□ (Opportunities) 글로벌 AI 학습 데이터의 고갈 위기와 버티컬 AI 시장의 급성장이 제공하는 전략적 도약 계기

- 글로벌 R&D 패러다임의 급격한 전환에 따른 데이터 주도형 혁신 수요 급증
 - 연구자의 직관과 반복 실험에 의존하던 기존 방식이 고비용·저효율의 한계에 봉착함에 따라, AI 기반 가설 설정 및 결과 예측을 통한 초고속·저비용 혁신 R&D가 기술 혁신의 뉴노멀로 부상함
 - 구글 딥마인드의 'GNoME'이 800년 분량의 신소재 구조를 단 수주 만에 발견하는 등 데이터 자산화를 통한 파괴적 혁신 사례가 창출되며 데이터 주도형 연구 체계에 대한 전 세계적 필요성 증대
- 범용 데이터 고갈 위기에 따른 고품질 전문 데이터(IP)의 전략적 가치 급상승
 - 고품질 AI 학습용 텍스트 데이터가 고갈될 것이라는 예측에 따라, 그 유일한 대안으로 특허와 같은 고신뢰 전문 데이터의 활용이 제시됨
 - AI 모델 성능 경쟁이 한계에 다다르며 고품질 데이터 확보가 AI의 성패를 가르는 데이터 중심 AI 시대로 진입함에 따라 희소성 높은 전문 데이터의 가치가 극대화됨
- 지식재산(IP) 데이터 자산 가치 부각
 - 특허 명세서 내 포함된 핵심 실험 데이터를 AI 학습용 자산으로 전환하기 위한 지능형 문서 처리

(IDP) 기술의 중요성이 부각되고 있으며, 비정형 데이터의 구조화를 통한 R&D 분석 효율성 제고에 대한 수요가 집중됨

- (기술 동향) 이미지·표 등 비정형 IP 데이터를 기계 독해가 가능한 형태로 정밀 분석하는 선도적 기술 표준화 움직임이 가속화되고 있으며, 이를 통한 'AI-Ready' 데이터 구축의 실질적 가능성을 입증함

○ 주요 선진국의 AI-Ready 데이터 인프라 구축 및 지능형 자산화 가속화

- 주요 선진국은 지식 자산이 AI 학습에 즉시 활용될 수 있도록 기계 독해가 가능한 표준 스키마 적용을 추진하고 있으며, 대규모 특허 정보를 지능형 R&D용 데이터셋으로 전환하여 민간에 개방하는 추세임
- 글로벌 특허 당국은 AI 기술을 활용한 정밀 분석 및 정보 추출을 연구 현장에 적극 도입하고, 표준화 데이터기반의 활용 생태계를 조성하여 R&D 전주기의 디지털 전환(DX)을 선도하고 있음

○ 글로벌 R&D 패러다임 전환에 따른 기계 가독성 및 지능형 분석의 필수 인프라화

- 데이터 주도형 R&D의 확산 및 지식 기반 혁신 경쟁은 데이터의 상호운용성 및 무결성 확보를 핵심 요건으로 설정하며, 고품질 IP 데이터 분석 기술의 도입을 R&D 성과 창출을 위한 필수 인프라로 정의함
- 국가 첨단전략산업의 R&D 생산성 혁신 및 초격차 기술 경쟁력 강화를 위해 특허 빅데이터와 AI를 융합한 지식재산 전주기 혁신 지원 필요성이 국가적 핵심 논의 사항이자 기회 요인으로 부상함

□ (Threats) 글로벌 기술 경쟁 심화 및 핵심 AI 플랫폼의 높은 해외 의존도

○ 글로벌 AI-Ready 표준화 경쟁 및 지능형 R&D 시장의 구조적 진입 장벽화

- 주요 경쟁국 주도의 AI 가독성 및 표준 스키마 선점이 글로벌 기술로 규범화 되어 표준 체계 미보유 시 12대 전략기술 등 첨단 분야의 생태계 진입 및 기술 협력이 원천적으로 제한되는 구조적 장벽이 강화되고 있음
- 국가 차원의 독자적인 지능형 IP 데이터 표준화 및 거버넌스를 조기에 확립하여 글로벌 기술 규범 변화에 선제적으로 대응하고, 국내 기업의 지능형 R&D 글로벌 시장 진출 및 협력 기반을 공고히 해야 함

○ 주요 선진국 중심의 기술 주권 강화 및 지식 자산 선점 경쟁 가속화

- 글로벌 경쟁국을 중심으로 특허 빅데이터의 AI-Ready 자산화 및 지능형 R&D 인프라 구축 전략이 가속화됨에 따라, 이에 대한 신속한 대응 실패 시 국가 차원의 AI 기술 주도권 및 R&D 경쟁력 약화 우려가 상존함
- 주요국의 기술 독점주의 정책 및 데이터 거버넌스 강화 추세에 대응하여, 12대 국가 전략기술 분야의 독자적인 지식 자산 확보와 데이터 표준화 체계 구축을 통한 기술적 우위 수호가 시급함

○ 핵심 AI 기술 및 플랫폼의 높은 해외 의존도로 인한 R&D 종속성 고조

- 지능형 분석 및 거대 모델 기술의 높은 대외 의존도로 인해, 글로벌 기술 환경 변화나 외부 플랫폼의 정책 변동 시 국내 R&D 생태계의 안정성 및 연속성이 저해될 수 있는 구조적 취약성을 내포함
- 세계 빅테크기업 중심의 지능형 서비스 독점과 데이터 장벽 강화에 대응하기 위해, 독자적인 IP-AI 핵심 알고리즘 및 자립적 생태계 확보를 통한 기술적 자립성 강화가 필수적임

□ SWOT 매트릭스 도출

표 8. SWOT 분석

		강점(S)	약점(W)
		내부 역량	외부 환경
기 회 (O)	○ 데이터 주도형 R&D 패러다임 전환에 따른 지능형 혁신 솔루션 수요 급증 ○ 학습 데이터 고갈 예고에 따른 고신뢰 전문 데이터 전략 가치 증대 ○ 해외 AI 자산화 사례 확산에 따른 국내 기술 벤치마킹 및 도입 여건 성숙 ○ 주요국의 AI-Ready 데이터 표준화 및 특허 데이터 전면 개방 경쟁 가속화	○ 5억 건 이상의 세계 특허 빅데이터를 국가전략 지식 자산으로 확보 ○ 특허 수명주기 및 권리 정보 연계를 통한 국가 차원의 분석 노하우 보유 ○ 관련 법령에 기초한 비정형 데이터 활용 및 상용화 법적 기반 확립 ○ 특허 실험 데이터 자동 추출 및 지능형 문서 처리 인프라 구축 역량 보유 ○ 세계 2위권 특허 경쟁력 기반의 고부가가치 기술 정보 자산화 역량 보유	○ 법률 용어 및 복합 이미지·표 형태에 따른 비정형 데이터 활용 장벽 존재 ○ 키워드 매칭 위주 기존 시스템의 AI 추론용 데이터셋 제공 한계 ○ 고난도 큐레이션 비용 및 기술 장벽에 따른 민간 주도 자산화 추진 정체 ○ AI 모델 오류 가능성에 따른 도메인 전문가의 데이터 검증 의존도 높음
		[SO 전략] - [SO1] 대량의 특허정보와 IDP 기술 결합으로 데이터 기반 처방형 R&D 패러다임 전환 주도 - [SO2] 12대 국가전략기술 분야의 고품질 특허 자산화로 글로벌 기술 경쟁력 확보 - [SO3] 관련 법령 기반 국가 IP 데이터 레이크 구축 및 데이터 주권 확립 - [SO4] 특허 내 실험데이터의 정밀 큐레이션으로 지능형 R&D 핵심 데이터 공급 체계 마련	[WO 전략] - [WO1] 비정형 정보 구조화로 단순 검색을 지능형 분석 환경으로 개편 - [WO2] 공공 데이터셋 개방으로 민간 AI 도입 비용 및 진입장벽 완화 - [WO3] 특허 전용 모델 개발로 R&D 현장 데이터 활용 장벽 제거 - [WO4] 실험 데이터와 공공 IP 융합으로 R&D 시행착오 단축 체계 확립
위 협 (T)	○ 국제 표준기구 중심의 사양 선점 경쟁에 따른 신규 기술 장벽 고착화 우려 ○ 글로벌 AI 기술을 활용한 경쟁국의 급격한 기술 추격 및 핵심 IP 권리 침해 위협 증대됨 ○ 글로벌 빅테크 API 생태계 독점에 따른 국내 지능형 서비스 기술 종속 위기 고조 ○ 기술의 국가종속 및 데이터 현지화 정책 강화에 따른 글로벌 공급망 리스크 확대	[ST 전략] - [ST1] 12대 전략기술 IP 자산화로 기술 패권 및 공급망 위기에 대응 - [ST2] 행정 역량 활용으로 AI 가독성 표준 제정 및 기술 장벽 무력화 - [ST3] 고도화된 추출 기술을 활용한 정밀 특허 분석으로 초격차 기술의 권리 방어를 지원 - [ST4] 고품질 데이터셋 자립화로 빅테크 종속성 완화 및 생태계 강화	[WT 전략] - [WT1] 독자적 IP-AI 알고리즘 개발로 해외 기술 의존 탈피 및 리스크 회피 - [WT2] 전문가 검증 기반 거버넌스 구축으로 AI 신뢰성 및 재현성 확보 - [WT3] 표준 데이터 스키마 기반의 범용성 확보로 특정 플랫폼 종속 및 기술 장벽 회피함 - [WT4] 클라우드 기반 구독형 서비스 보급으로 기술 도입 지속성 확보

표 9. SWOT 시사점

SWOT 시사점	전략	추진 전략
○ 국가 전략 지식 자산 활용을 통한 데이터 주도형 R&D 선도 - 5억 건의 특허 자산과 IDP 기술을 결합하여 기존 경험 중심 연구를 데이터 기반 처방형 R&D로 전환하고, 12대 전략기술 중심의 자산화로 글로벌 기술적 우위를 점해야 함 - 명확한 법적 근거를 바탕으로 국가 IP 데이터 스페이스를 구축하여 데이터 주권을 확립하고, 특허 내 실험데이터 정제를 통해 지능형 R&D의 핵심 동력을 상시 공급해야 함	- [SO1] 대량의 특허정보와 IDP 기술 결합으로 데이터 기반 처방형 R&D 패러다임 전환 주도 - [SO2] 12대 국가전략기술 분야의 고품질 특허 자산화로 글로벌 기술 경쟁력 확보 - [SO3] 관련 법령 기반 국가 IP 데이터 스페이스 구축 및 데이터 주권 확립 - [SO4] 특허 내 실험데이터의 정밀 큐레이션으로 지능형 R&D 핵심 데이터 공급 체계 마련	전략적 IP 데이터 자산화 체계 확립 ⇕ AI 기반 지능형 R&D 혁신 플랫폼 고도화 ⇕ 기술 가치사슬 확장 및 실증 가속화
○ 국가 차원의 역량 결집을 통한 기술 주권 및 글로벌 기술 패권 대응 - 12대 국가전략기술의 IP 자산화로 공급망 위기에 선제 대응하고, 행정 역량 기반의 AI 가독성 글로벌 표준 제정을 주도하여 대외 기술 장벽을 선제적으로 극복해야 함 - 고도화된 데이터 추출 기술로 초격차 기술의 권리 방어 역량을 제고하고, 고품질 데이터셋 자립화를 통해 글로벌 빅테크 종속성을 탈피하여 독자적 지능형 R&D 생태계를 구축해야 함	- [ST1] 12대 전략기술 IP 자산화로 기술 패권 및 공급망 위기에 대응 - [ST2] 행정 역량 활용으로 AI 가독성 표준 제정 및 기술 장벽 무력화 - [ST3] 고도화된 추출 기술을 활용한 정밀 특허 분석으로 초격차 기술의 권리 방어를 지원 - [ST4] 고품질 데이터셋 자립화로 빅테크 종속성 완화 및 생태계 강화	
○ 인프라 고도화 및 공공 주도의 생태계 활성화로 혁신 가속 - 비정형 정보의 구조화를 통해 기존 단순 검색 인프라를 지능형 분석 환경으로 개편하고, 특허 전용 모델 개발로 R&D 현장의 데이터 활용 장벽을 근본적으로 제거해야 함 - 공공 주도의 데이터셋 개방으로 민간 AI 도입 비용 및 진입장벽을 획기적으로 완화하며, 파편화된 실험 데이터와의 융합을 통해 R&D 전주기 시행착오를 단축하는 자산 체계를 확립해야 함	- [WO1] 비정형 정보 구조화로 단순 검색을 지능형 분석 환경으로 개편 - [WO2] 공공 데이터셋 개방으로 민간 AI 도입 비용 및 진입장벽 완화 - [WO3] 특허 전용 모델 개발로 R&D 현장 데이터 활용 장벽 제거 - [WO4] 실험 데이터와 공공 IP 융합으로 R&D 시행착오 단축 체계 확립	

○ 기술적 자립화 및 표준 거버넌스를 통한 R&D 지속성 확보

- 해외 의존도가 높은 아키텍처 대신 독자적인 IP-AI 핵심 알고리즘을 개발하고, 전문가 검증 기반의 거버넌스를 구축하여 AI 모델의 신뢰성과 연구 재현성을 극대화해야 함
- 표준 데이터 스키마를 통해 특정 플랫폼 종속 리스크를 회피하고, 중소기업 수용성을 고려한 클라우드 기반 구독형(SaaS) 서비스를 보급하여 기술 도입의 지속성을 보장해야 함

- [WT1] 독자적 IP-AI 알고리즘 개발로 해외 기술 의존 탈피 및 리스크 회피

- [WT2] 전문가 검증 기반 거버넌스 구축으로 AI 신뢰성 및 재현성 확보

- [WT3] 표준 데이터 스키마 기반의 범용성 확보로 특정 플랫폼 종속 및 기술 장벽 회피

- [WT4] 클라우드 기반 구독형 서비스 보급으로 기술 도입 지속성 확보

제2절 비전 및 목표

- 대내외 환경분석 및 SWOT 분석 결과, 지식재산 데이터 큐레이션 R&D는 비정형 IP 데이터의 고정밀 구조화 및 표준화를 통해 고부가가치 연구 데이터 자산 체계 확립이 필요한 것으로 도출
- 본 사업의 비전은 ‘비정형 IP 데이터의 지능적 자산화를 통한 R&D 패러다임 대전환 및 국가 전략기술 초격차 선점’이며, 기대효과는 ‘국가 R&D 생산성 극대화와 AI기반 지능형 R&D 생태계 활성화’임

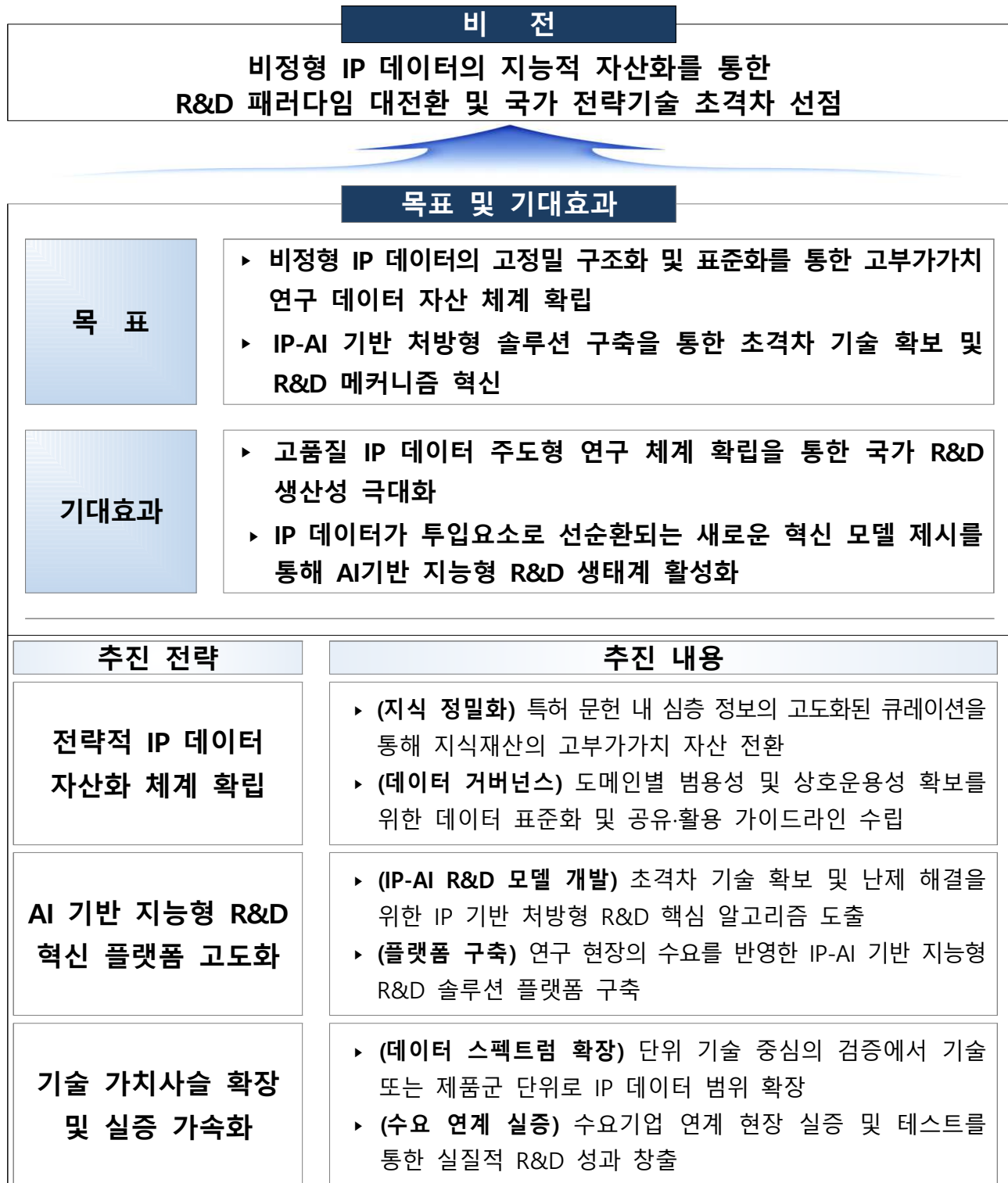


그림 21. 비전 및 목표

제3절 핵심 추진 과제

□ 지원 범위 : 12대 국가전략기술 분야

- 12대 국가전략기술, 특히 구조식, 수식, 도면 등 비정형 데이터가 상대적으로 많고, R&D 난제 해결을 위해 특허 데이터 기반의 수요가 많은 분야 집중 지원
- 12대 국가전략기술(50개 세부기술) 기준 데이터 기반 R&D 도입 시 시행착오 감축 효과가 큰 분야를 편셋 타겟팅하여 집중 지원
- 국가전략분야 50개 중점기술을 대상으로 비정형 데이터양 및 밀집도, 특허 기반 연구개발 전략(IP-R&D) 특허 수 등의 정량 통계 기준과 전문가 평가를 통해 우선 지원 순위 도출(반도체·디스플레이, 이차전지, 첨단바이오 등)

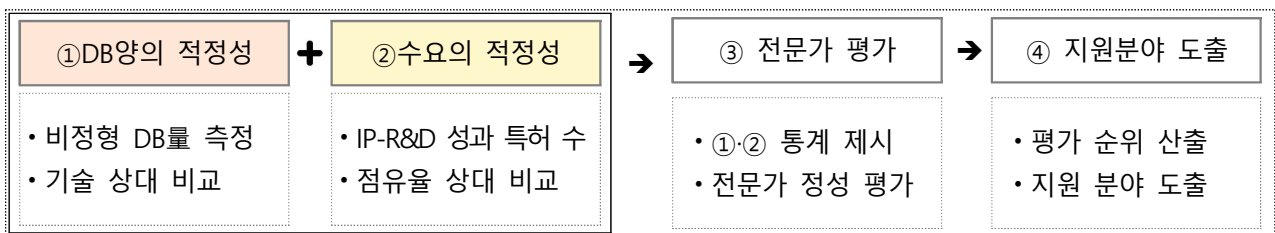


그림 22. 국가전략기술분야의 검증 가능 테마 도출 프로세스

- (①비정형 DB 量) 중점기술별 특허문헌 내 포함된 비정형 데이터(구조식, 표, 도면 등)의 양을 측정하고, 기술 간 상대 비교 도출
 - (②R&D 수요 실적) 특허로 R&D 문제를 해결하는 'IP-R&D 전략지원사업'의 특허 성과를 기술별로 측정하여 수요 집중 분야 도출
 - (③전문가 평가) 기술수요조사 결과와 ①비정형 DB의 적정성과 ②R&D 수요 적정성의 결과를 전문가에게 제시하고, 최종 전문가 판단*하여 순위 도출
- * (기준) 다양성(DB양), 파급성(DB 활용), 가능성(검증 효율성(PoC)), 유용성(중요성)

표 10. 특허문헌 내 비정형 데이터 분석·활용 요건

구분	내용
다양성	문헌내 비정형 데이터 다양성(수식, 구조, 이미지, 공정 조건 등)
파급성	산업계에서 실제로 활용하고자 하는 수요(R&D, 설계, 소재 탐색 등)
가능성	문헌내 비정형 데이터의 LLM/AI로 추출·정규화·분석 가능성
유용성	분석 결과에 대한 의사결정 활용 유용성

- (④지원분야 도출) 전문가 평가 결과에 따른 기술별 우선 순위 부여, 우선 순위가 높은 중점기술로 구성된 기술분야 도출

□ 지원 대상

- (대상개요) R&D를 위해 맞춤형 AI학습용 특허 데이터셋 구축을 추진하고자 하는 수요가 존재하는 산·학·연 연구개발기관 및 데이터 큐레이션 기관

* IP 데이터를 R&D의 결과물이 아닌 핵심 생산·투입요소로 활용하며, 비정형 IP 데이터의 고정밀 큐레이션 역량 및 R&D 혁신 파급효과 등을 고려하여 분야별 기술난제 해결이 가능한 컨소시엄을 선별함

○ (수행체계) 산·학·연 등 데이터셋 수요 기관과 데이터 큐레이션 기업 등이 공동 참여하는 개방형 컨소시엄

- 수요 산·학·연(실증, 주관), 특허 서비스·데이터 전문기관(큐레이션, 공동), AI·SW 전문기업(플랫폼 개발, 공동) 등 유기적 협력 컨소시엄 구성

* 현장 실증 역량을 갖춘 연구개발 수요가 존재하는 대학·출연연·기업(실증기관)을 주관으로 도메인 특화 AI 기술력을 보유한 기업(AI·SW 전문기업), 그리고 IP 데이터 자산화 전문성을 가진 큐레이션 기관(특허 서비스·데이터 전문기관) 간의 유기적 협력 컨소시엄 구성을 필수 요건으로 함

표 11. 컨소시엄 구성 예시

수요 기관 (주관기관) ⇔ 데이터 큐레이션 기관 ⇔ AI·SW 기업				
산·학·연		IP 데이터 서비스 기관 빅데이터 연구 기관		플랫폼 개발 기관
구분	주관연구개발 기관(실증중심)	공동연구개발 기관1	공동연구개발 기관2	컨소시엄별 특징
표준형 컨소시엄(기업 주도)	중소·중견 제조기업 (반도체, 소재 등)	AI·SW 전문기업	특허법인/ 특허사무소	IP-AI를 활용한 산업혁신 주도
전략기술 표준화 컨소시엄 (공공/대학 주도)	출연연, 대학 또는 기술집약형 기업 (기술이전 대상자)	IP서비스 전문기관 (IP-AI 융합 스타트업 등 포함)		국가 전략기술 난제 해결 및 공공 표준화

※ 컨소시엄 구성 상세 예시

① (실증 중심 표준형 컨소시엄) IP-AI를 활용한 산업혁신 주도

- (컨소시엄 목적) 제조 및 기술개발 현장의 수요를 반영한 고품질 특허 데이터를 AI 전문기업이 학습하여, 설계 시간 단축 및 실험 비용 절감 등 제조공정 및 연구개발 전반에 즉시 적용 가능한 'R&D 최적화 지능형 솔루션' 구축을 목표로 함
- 협력 구조
 - 중소·중견 제조기업(주관/실증) : 반도체나 신소재 분야의 R&D 과정에 솔루션을 도입하여, 설계 시간 단축 및 실험비 감소 등을 현장에서 직접 검증함
 - AI·SW 전문기업(공동1) : 특허 비정형 데이터 추출 및 기술 추론을 위한 코어 알고리즘을 개발하고, 사용자 친화적인 솔루션 인터페이스(UI/UX)를 구축함
 - 특허법인(공동2) : 수십만 건의 기술 명세서 및 도면을 AI가 학습 가능한 기계 가독성 데이터로 정밀 큐레이션하여 엔진의 분석 정확도와 법적 신뢰성을 확보함
- (예상 성과) 특정 산업 및 개별 기업의 R&D 특성에 최적화된 맞춤형 솔루션을 보급함으로써, 민간 주도의 데이터 기반 기술 혁신 생태계 조성 및 산업 경쟁력 강화

② (전략기술 고도화 컨소시엄) 산·학·연 협력을 통한 전략기술 난제 해결 및 기술 고도화

- (컨소시엄 목적) 대학·출연연의 원천 기술력과 기업의 현장 요구사항을 결합하여, 국가 전략기술 분야의 핵심 기술 난제를 공동으로 해결하고 참여 기업의 직접적인 기술 수준 성장 및 고도화를 목표로 함
- 협력 구조
 - 출연연, 대학 또는 기술집약형 기업 (주관/실증중심): 원천기술의 심층 분석과 고난도 알고리즘 설계를 주도하며, 기술집약형 기업과 공동 R&D를 통해 현장의 기술적 난제에 대한 실질적인 해결 방안을 마련
 - IP서비스 전문기관 (공동1): IP-AI 융합 스타트업 등이 참여하여 대규모 비정형 데이터를 분석하고, 난제 해결을 위한 처방형 R&D 로직을 플랫폼상에서 구현하여 연구 효율을 극대화함
 - 특허법인/특허사무소 (공동2): 기술 명세서 및 도면의 데이터 큐레이션을 통해 엔진의 정확도를 높이는 동시에 해결된 기술 난제에 대한 핵심 특허 확보 및 법적 보호 전략을 수립함
- (예상 성과) 데이터 주도형 산학연 공동 R&D 지원을 통해 국가 전략기술의 난제를 실질적으로 해결하고, 기술집약형 기업의 초격차 기술 경쟁력을 조기에 확보함

□ 지원기준

- (지원단계) 기술사업화를 위한 기술개발, 제품제작 등 사업화 단계를 지원하며, 실증 대상 기술이 TRL4 단계 이상에 해당하는 연구개발만 지원 가능

구분	TRL1	TRL2	TRL3	TRL4	TRL5	TRL6	TRL7	TRL8	TRL9
일반	기본원리	개념정립	개념검증	시작품		시제품	실용화		양산
바이오	기본원리	개념정립	개념검증	in vivo 검증	GLP 검증	임상1상	임상2상	임상3상	사업화
내용	기초이론 정립	기술개념 정립	특허출원/기본성능 검증	설계제작	성능평가 / 비임상	설계제작 성능평가	신뢰성/안정성/유효성 평가	시제품 인증 / 인허가	시판/시판 후 연구

그림 23. TRL 예시

- (TRL단계 참고) 이론적으로 정립된 기술 개념을 실제 실험실 규모의 시연 모델(Mock-up)이나 시제품(Prototype)으로 구현하여 기본적 기능 작동성이 확인된 상태로써 연구개발 과정에서 해결해야 하는 기술적 난제가 명확히 식별되어 있으며, 이를 해결하기 위한 기초 데이터가 확보된 단계임

* (의약품/바이오) 신규 타겟에 대한 작용 기전(MoA)이 규명되고, 실험실 수준(In-vitro)에서 유효성을 보이는 선도물질 또는 후보물질이 도출되어 기초 실험 성적서가 확보된 단계

* (반도체/디스플레이) 설계한 회로 혹은 소재의 특성이 전용 설계 툴(EDA)이나 시뮬레이터 상에서 목표로 하는 임계치(성능, 수율 등)를 만족하며, 물리적 구현을 위한 설계 최적화 직전 단계

* (소재/부품) 신규 소재의 조성비나 부품의 구조가 연구실 단위의 샘플(Coupon) 제작을 통해 목표 물성을 구현함을 확인하고, 대조군 대비 기술적 우위가 실험적으로 증명된 상태임

□ 단계별 지원내용

- **(기본방향)** 연구개발의 효율화를 위해 특허 데이터를 AI 학습용 고품질 데이터셋으로 제공하고, 이를 기반으로 하는 연구개발 솔루션 지원
- **(차별성)** 범용 데이터셋 중심의 타 부처 사업과 달리, 지식재산(IP) 데이터를 단순 결과물이 아닌 연구개발 핵심 생산·투입 요소로 재정의하여 신기술 개발의 시행착오를 획기적으로 단축하는 R&D 패러다임 전환
- **(전체 지원내용)** 특허 내 비정형 데이터를 정밀 큐레이션(자산화)하고, 이를 기반으로 최적의 R&D 경로를 제시하는 지능형 R&D 실증 수행
 - (1단계) 연구개발 기술 진단 및 내부의 연구 데이터와 외부의 빅데이터(특허, 논문 등)를 정제하여 AI학습에 최적화된 고품질의 데이터(실험조건, 조성물 배합 비율, 성능 수치, 구조식, 문제·해결 수단, 도면 등) 구축
 - (2단계) 연구개발 맞춤형 AI모델 구축, AI활용 연구설계 등을 통해 최적의 R&D 방향 도출하고, 가상실험 등으로 신규 아이디어 도출
 - (2-1) 지식재산(IP) 기반 인공지능(AI) 활용 솔루션(IP-AX) 모델을 활용하여 연구설계 및 가상실험 등을 실시하고 최적화된 아이디어 등 아이템 도출
 - (2-2) 도출된 아이디어에 대한 산업 현장의 물리적 연구·실험을 통해 신규 기술 확보 및 특허 권리화를 지원하고, IP-AX 제품 개발

구분	1단계(1년)	2단계(2년)	
	R&D 진단 및 고품질 학습데이터 확보	맞춤형 IP-AX 모델 개발 및 가상 실험 검증	연구 현장 실증 및 제품 개발 혁신
내용	• R&D 진단·방법론 수립 • 연구 DB 분석·표준화 • 논문·특허 DB 수집·가공 • 학습 DB 구축 등 • PoC 검증	• 연구 범위 및 DB 확장 • AI 모델 커스터마이징 • R&D 연구 설계 • 가상 실험·검증 • 아이디어 최적화	• 아이디어 현장 실증 • 신규 기술 검증·확보 • 핵심 특허 권리화 • IP-AX 모델 최적화 • 제품 개발 및 사업화
산출 성과	고품질 AI 학습 DB 셋	IP-AX 솔루션 모델	핵심특허/IP-AX제품

그림 24. IP-AX 제품 개발 단계

□ 단계별 세부 지원 내용

- **(1단계)** 비정형 특허 데이터 큐레이션 및 핵심 알고리즘 유효성 검증
 - (IP 데이터 큐레이션) 특허 내 비정형 데이터(마쿠쉬 구조, 실험 테이블, 수치 범위 등)를 정밀 추출하여 정형화·표준화된 데이터 및 AI 학습 가능 고부가가치 자산 구축
 - * 연구 현장의 수요를 반영하여 기술 도메인별 핵심 파라미터(온도, 압력, 배합비 등) 중심의 딥 큐레이션 수행
 - (핵심 알고리즘 기반 개념실증(PoC)) 큐레이션 데이터 기반 핵심 알고리즘 수립 및 Lab 수준의 실험·시뮬레이션을 통한 기술 유효성 조기 검증

- (데이터 플랫폼 기반 확립) 분야별 데이터 거버넌스를 수립하고, IP 특화 데이터의 수집·적재 프로세스 및 고품질 데이터 유지를 위한 품질관리(QA) 기준 확립
- (2단계) 데이터 스펙트럼 확장 및 솔루션 고도화를 통해 현장 실증 및 상용화
 - (데이터 범위 확장) 기술 또는 제품군 단위 IP 데이터 레이크(Data Lake) 구축 및 전·후방 연계 기술군 데이터 클러스터링을 통한 통합 자산화 수행
 - (데이터 정제·연계) 고밀도 융합형 데이터셋의 상시적·자동적 자산화를 위한 정제 SW를 개발하고, 신규 유입된 데이터의 즉각적인 수용과 확장이 가능한 범용 표준 테이블 및 기술 상관관계 모델링 구축
 - * 새로운 특허 및 실험 데이터 도출 시, 이를 지속적으로 추출하고 기존 지식 스키마에 자동 통합함으로써 연구개발 전 주기에 걸친 지능형 자산 업데이트 체계 마련
 - (소규모 성능 검증 및 미세조정) 대규모 실증 전, 제한된 범위의 실제 산업 데이터 또는 소규모 테스트베드(Bench-scale)를 활용하여 처방형 AI 모델의 지속적인 파인튜닝 및 알고리즘 최적화 수행
 - (솔루션 고도화 및 아키텍처 설계) 전·후방 통합 데이터를 수용하기 위한 대규모 통합 아키텍처를 설계하고, 파인튜닝된 AI 모델과 솔루션 간의 유기적 연계 구현
 - (현장 실증 및 제품 개발 검증) 수요기업 연계 파일럿 테스트를 통한 기술 분야별(치료제, 반도체 등) 실질적 제품 개발 성과 도출 및 유효성 확인
 - * 실증 사이트 내 IP-AX 솔루션 적용 전·후의 제품 개발 데이터 비교 및 기술적 목표 달성도 정밀 검증
 - * 현장 실시간 데이터 피드백 기능을 활용한 제품 개발 모델의 파라미터 최적화 및 결과물 신뢰성 확보
 - (R&D 공정 및 모델 최적화) 실증 과정에서 도출된 현장 데이터를 반영하여 알고리즘의 정밀도를 극대화하고, 실제 R&D 과정에 최적화된 고성능 연산 환경 및 표준화 체계 구현
 - * 실증 데이터 기반의 모델 파인튜닝으로 예측 정밀도를 제고하고 대규모 R&D 데이터 처리에 최적화된 고성능 컴퓨팅 연동 환경 구축
 - * 현장 대응형 모델 경량화 및 표준 데이터 포맷 개발을 통해 연산 신속성을 확보하고 기술 도메인 간 데이터 호환성 체계화
 - (R&D 신뢰성 제고 및 방법론 유효성 분석) 연구 전주기의 데이터 무결성을 유지하고, 개발된 IP-AX 주도형 R&D 방법론의 기술적·경제적 타당성 검증
 - * 연구 데이터 이력 추적 기반의 투명성을 확보하여 연구 결과의 무결성 중심인 지능형 통합 연구 관리 체계 고도화
 - * 기존 방식 대비 기간 및 비용 절감 효과를 정량적으로 분석하여 신규 방법론의 타당성 입증 및 분야별 최적 활용 가이드라인 도출

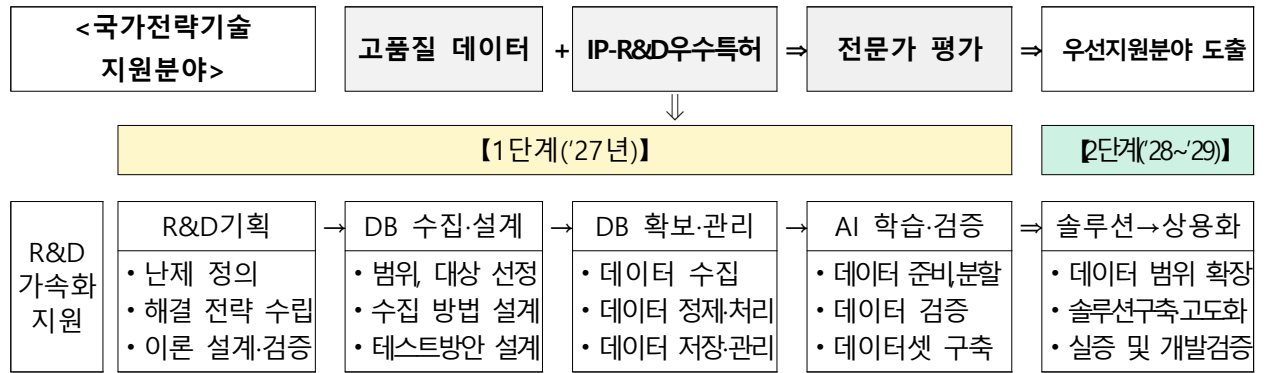


그림 25. 우선지원분야 도출 계획

□ 세부지원 분야

- (추진방향) 특허 분석자료 등을 기반으로 추진 로드맵 및 후보 기술군을 도출하고, 이를 기초로 세부 기술테마 확정 및 상세 목표/요구사항 등 도출
- 세부기술분야는 반도체/디스플레이, 이차전지, 첨단바이오, ICT표준(차세대통신, AI 및 첨단로봇·제조)을 선정

표 12. 지식재산 데이터 큐레이션 중점 지원 분야

전략기술분야		중점 지원분야	
		기술명	지원 분야 수
반도체·디스플레이		①고성능·저전력 AI 반도체(NPU 등), ②차세대고성능센서 ③소재·부품·장비, ④반도체 첨단 패키징	4
이차전지		①리튬이온전지 및 핵심소재, ②차세대 이차전지 소재·셀, ③이차전지 모듈·시스템, ④이차전지 재사용·재활용	4
첨단 바이오		①합성생물학, ②유전자·세포치료, ③감염병 백신·치료, ④디지털헬스데이터 분석·활용	4
ICT 표준	차세대 통신	①5G Adv-6G 이동통신, ②Open-RAN ③고효율 5G-6G 통신부품	3
	인공지능	①첨단 AI 모델링·의사결정, ②산업 활용·혁신 AI	2
	첨단로봇·제조	①로봇 정밀제어·구동 부품·SW, ②고난도 자율조작, ③인간·로봇 상호작용	3
합계			20

○ 반도체/디스플레이

표 13. 반도체/디스플레이 분야 과제 개요

구분	상세 내용
기술개요	본 기술은 초미세 공정 한계 극복을 위한 고성능·저전력 AI 반도체 설계와 칩 간의 데이터 전송 효율을 극대화하는 첨단 패키징(HBM, 3D 적층 등), 그리고 이를 구현하기 위한 핵심 소부장(소재·부품·장비) 기술을 포함함
필요성	(기술 패권 및 데이터 패쇄성 대응) NVIDIA·TSMC 중심의 HBM 등 첨단 패키징 독점에 맞서, 기업비밀로 관리되는 제조 레시피 대신 공개 특허에서 핵심 지식을 추출하는 전략이 시급함 (TRL 단계 이행 및 변수 제어) TRL 변화 등 이행 시 발생하는 CTE(열팽창계수), 범프 간격 등 비정형 공정 변수를 제어하기 위해 AI 기반 자동화 큐레이션 체계 도입이 필수적임 (역공학을 통한 기술 주권 확보) 데이터 주도형 역공학으로 성공 가능성이 검증된 설계도를 선제 확보하여, 과거 소부장 국산화 경험을 지능화된 솔루션으로 자산화하고 국가 디지털 주권을 수호함
범위 및 목적	(최종 목표) 차세대 반도체 및 첨단 패키징 분야의 전 세계 특허 빅데이터를 자산화하여, TRL 4단계 시작품의 성능 예측 정밀도를 높이고 공정 수율을 20% 이상 개선할 수 있는 지능형 R&D 솔루션 개발 (지원 범위) TRL 4단계(실험실 수준의 시운전 및 시작품 제작) 이상의 상용화 연계 기술 개발 및 실증 데이터셋 구축 (대상 분야) 고성능·저전력 AI 반도체(NPU 등), 차세대 고성능 센서, 반도체·디스플레이 소재/부품/장비, 첨단 패키징(HBM, CoWoS 등)
세부 기술 요구사항	① 고성능·저전력 AI 반도체 (NPU, PIM 등) - 데이터 추출 요구사항 - 특허 명세서 내 Power, Performance, Area 핵심 지표 및 아키텍처 데이터 구조화 - 회로도 이미지에서 로직 게이트 배치 및 메모리 셀 구조(SRAM/MRAM 등) 수치 추출 - 실증 내용 - 특허 데이터로 학습된 예측 모델을 통해 산출된 시작품의 전력 효율과 실제 FPGA/Test Chip 측정값 간의 오차율 10% 이내 검증 - 동적 전압·주파수 조절 알고리즘 적용 시의 전력 저감 효과를 실측 데이터로 증명함 ② 차세대 고성능 센서 (LiDAR, 고해상도 이미지 센서 등) - 데이터 추출 요구사항 - 픽셀 구조별 광전 변환 효율, 신호 대 잡음비, 양자 효율 등 성능 데이터 큐레이션 - 렌즈 및 필터의 광학적 특성(투과율, 굴절률) 및 파장 대역별 응답 특성 수치 추출 - 실증 내용 - 가혹 환경(85°C 고온 및 85% 고습) 조건에서 센서의 해상도 및 검출 거리 유지 성능을 데이터로 검증함 - 특허 기반으로 도출된 최적 픽셀 설계안을 적용한 시작품의 SNR 개선 수치가 기존 대비 15% 이상 향상됨을 실증함 ③ 반도체·디스플레이 소재·부품·장비 (소부장) - 데이터 추출 요구사항 - 감광액, 전구체 등 핵심 소재의 화학적 조성비, 분자량, 열적 안정성 데이터 큐레이션 - 장비별 챔버 압력, 플라즈마 파워, 가스 유량 등 실제 제조 레시피의 수직 계열화 구축

	• 실증 내용 - 구축된 데이터셋을 적용하여 제작된 소자 시작품의 식각 및 증착 균일도가 95% 이상 달성됨을 확인 - 국산화 부품 적용 시 장비의 평균 고장 간격(MTBF)이 특히 데이터 기반 예측치와 부합함을 시운전을 통해 입증함 ④ 반도체 첨단 패키징 (HBM, 2.5D/3D 패키징 등) • 데이터 추출 요구사항 - TSV 피치, 범프 크기, 언더필 소재의 점도 및 탄성 계수 추출 - 다층 적층 구조에서의 열 저항 및 신호 무결성 변수 큐레이션 • 실증 내용 - 적층 단수 증가에 따른 휨 현상 제어 수치가 특히 데이터 기반 시뮬레이션 결과 대비 5% 이내의 정밀도를 가짐을 실측함 - 시작품의 열 방산 효율이 글로벌 특허 규격과 비교하여 동등 수준 이상의 방열 성능을 확보했음을 데이터로 증명함
--	---

도면2

소재구성

반도체는 소수성 유기반도체를 다양한 용매에 녹여 다양한 농도의 용액으로 제조한 후, 이 용액을 관중을 위해 코팅하여 형성할 수 있다. 용액은 산도로부터, 알도로부터, 디알도로부터 등을 사용할 수 있다. 용액 중 유기반도체의 농도는 5 내지 10 mg/ml일 수 있다. 용액에 유기반도체를 완전히 녹이기 위해, 90 내지 110℃ 및 1 내지 3시간 조건으로 가열 교반할 수 있다. 코팅방법으로는 스핀 코팅방법 등을 이용할 수 있다. 스핀 코팅의 속도 및 시간은 각각 2000 내지 3000 rpm 및 60 내지 120초일 수 있다. 소수성 유기반도체 트랜지스터의 전기적 성능 향상을 위해 코팅 후 열처리를 할 수 있으며, 열처리 온도 및 시간은 각각 150 내지 200℃ 및 20 내지 40 분일 수 있다. 활성층의 두께는 20 내지 50 nm일 수 있다. 활성층의 두께가 너무 두꺼우면 표면거칠기가 적게 형성될 수 있다.

반도체는 활성층과 기판 사이에 형성될 수 있고, 소수성 유기반도체의 분자들간 응집 관련 접착에너지를 유도할 수 있으며, 소수성 유기반도체의 분자들간 응집에너지도

3-옥솔[Poly(1,4,5,7-di-fluoromethyl-2,6-di-fluoro-3-oxole): PFD, Asahi Glass에서 상표명 Cytop으로 상업적으로 판매되고 있음], 폴리디메틸실록산(Polydimethylsiloxane: PDMS), 폴리스티렌(Polystyrene: PS), 폴리메틸메타크릴레이트(Polyethylmethacrylate: PMMA), 폴리(비닐알코올)(Poly(vinyl alcohol): PVA), 폴리(4-비닐페닐)(Poly(4-vinyl phenyl): PVP), 옥틸트리클로로실란(Octyltrichlorosilane: OTS), 옥타데실트리클로로실란(Octadecyltrichlorosilane: OTS) 등을 단독으로 또는 2종 이상 혼합하여 사용할 수 있다. 배반적계는, 표면에너지가 작은 PFD(Cytop)를 사용할 수 있다.

반도체는 저 표면에너지 소재를 기판 위에 코팅하여 형성할 수 있다. 코팅 전에 기판을 전-처리할 수 있다. 전-처리 방법으로는 산소-플라즈마 처리 등을 이용할 수 있고, 처리시간은 1 내지 3분일 수 있다. 코팅방법으로는 스핀 코팅방법 등을 이용할 수 있다. 스핀 코팅의 속도 및 시간은 각각 3000 내지 4000 rpm 및 40 내지 60초일 수 있다. 코팅 후 110 내지 130℃에서 10 내지 30분간 가열할 수 있다. 활성층의 두께는 0.1 내지 10 nm일 수 있다. 활성층의 두께가 너무 두꺼우면 트랜지스터 성능이 저하될 수 있다.

반응조건

반도체 표면에 나노스케일의 거칠기를 극대화하기 위하여, 본 발명에서는 유기반도체 활성층-기판간 접착에너지의 차이를 제어하는 방법을 사용한다.

소수성 유기반도체의 분자들간 응집에너지는 65 mJ/N 이상일 수 있다. 응집에너지의 상한치는 예를 들어 100, 90, 80 또는 70 mJ/N일 수 있다. 응집에너지는 접촉각 측정기, 표면장력계 등을 이용하여 측정할 수 있다.

재료특성

반응시간

도면3

구조식

그림 26. 반도체 큐레이션 대상 항목 예시

○ 이차전지

표 14. 이차전지 분야 과제 개요

구분	상세 내용
기술개요	본 기술은 고에너지 밀도와 안전성을 동시에 확보하기 위한 리튬이온전지 고도화 및 화재 위험을 근본적으로 제거한 차세대 이차전지(전고체 등) 개발, 그리고 배터리 생애주기 전반의 모듈 시스템 및 재활용 기술을 포괄함
필요성	• (공급망 재편 및 차세대 기술 주권) 전고체 등 차세대 이차전지 시장 선점을 위한 고에너지 밀도 소재 확보와 배터리 여권 등 글로벌 규제에 대응하는 공신력 있는 데이터 자산 구축이 시급함

	• (공정 복잡성 및 병목 구간 해소) TRL 변화 이행 시 발생하는 배합비, 코팅 두께 등 수천 가지 공정 변수 간 인과관계 규명의 어려움을 극복하여 개발 지연 문제를 해결해야 함 • (데이터 주도형 R&D 혁신) 기존 시행착오 방식에서 벗어나 특허 내 방대한 실험·공정 데이터를 정밀 큐레이션함으로써 최적화된 기술 지침을 도출하고 국가 R&D 생산성을 혁신함
범위 및 목적	• (최종 목표) 이차전지 전주기 특허 빅데이터를 자산화하여, 이차전지 성능(에너지 밀도, 수명 등) 예측 정밀도를 높이고 제조 공정 효율을 20% 이상 개선 • (지원 범위) TRL 4단계(실험실 수준의 시운전 및 시작품 제작) 이상의 상용화 연계 기술 개발 및 실증 데이터셋 구축 • (대상 분야) 리튬이온전지 및 핵심소재, 차세대 이차전지 소재·셀, 이차전지 모듈·시스템, 이차전지 재사용·재활용
세부 기술 요구사항	① 리튬이온전지 및 핵심소재 (양극·음극재, 전해질, 분리막) • 데이터 추출 요구사항 - 하이닉셀(High-Ni) 양극재 및 실리콘(Si) 음극재의 화학적 조성비, 입자 크기(D50), 표면 코팅층 두께 데이터 큐레이션 - 전해질 첨가제 조성에 따른 전해액-전극 계면(SEI) 형성 메커니즘 및 열적 안정성 수치 추출 • 실증 내용 - 특허 기반으로 도출된 최적 소재 배합비를 적용한 0.5Ah~1Ah급 파우치/원통형의 초기 방전 용량 및 쿨롱 효율 측정 - 소재 특성 데이터 기반 예측 모델을 통한 수명 예측치와 실제 100회 이상 충·방전 사이클 테스트 결과 간의 오차율 10% 이내 검증 ② 차세대 이차전지 소재·셀 (전고체, 리튬황, 전고체 전해질 등) • 데이터 추출 요구사항 - 고체 전해질의 이온 전도도(Ionic Conductivity), 계면 저항 수치 및 소재별 기계적 강도(탄성 계수 등) 데이터 구조화 - 리튬 금속 음극의 덴드라이트(Dendrite) 억제 기술 및 전극-전해질 접합 공정 변수 큐레이션 • 실증 내용 - TRL 4단계 전고체 전지 시작품의 고온(60도) 및 저온(20도) 구동 특성이 특허 내 선행 기술 지표 대비 동등 이상임을 입증 - 적층(Stacking) 공정 시의 가압 조건 변화에 따른 셀 내부 저항 감소 효과를 실측 데이터로 증명함 ③ 이차전지 모듈·시스템 (BMS, 열관리 시스템 등) • 데이터 추출 요구사항 - 셀 간 전압 편차 제어 알고리즘, 충전 상태(SOC) 및 수명 상태(SOH) 추정 알고리즘의 비정형 수식 데이터 큐레이션 - 모듈 냉각 방식(수랭/공랭)에 따른 온도 분포 수치 및 열폭주 전이(Thermal Runaway Propagation) 차단 구조 데이터셋 구축 • 실증 내용 - 배터리 관리 시스템(BMS)이 특정 이상 상황(과충전, 단락) 발생 시 특허 데이터 기반 예측 시나리오대로 안전 기능을 수행하는지 실증함 - 고출력 방전 시 모듈 내부의 최고 온도가 시뮬레이션 결과 대비 5도 이내의 정밀도로 관리됨을 확인

	④ 이차전지 재사용-재활용 (진단, 유가금속 회수) • 데이터 추출 요구사항 - 폐배터리의 잔존 가치 평가를 위한 비파괴 검사 지표 및 용량 퇴화 패턴 데이터 큐레이션 - 습식 제련 공정 내 용매 추출 단계별 유가금속(리튬, 니켈, 코발트 등) 회수율 및 불순물 제거 효율 수치 추출 • 실증 내용 - 재활용 공정을 통해 회수된 소재를 재사용한 시작품(Regenerated Cell)의 성능이 신규 소재 대비 95% 이상 유지됨을 실증 데이터로 입증함 - 특허 기반의 최적 회수 공정 적용 시 유가금속의 최종 순도가 국제 표준 규격을 만족함을 확인
--	--

[0059] **실시예 1: 흑연질 탄소결합물 (G/CN)의 합성**

[0060] 6 g의 멜라민(Melamine) 분말을 1000 mL 증류수가 담긴 비커에 넣고, 100 ℃로 설정된 핫플레이트 위에서 멜라민이 완전히 녹을 때까지 60분 동안 교반하였다. 그 다음 60% 농도의 질산 용액을 5 mL 넣고, 용액이 모두 증발하여 건조한 분말 형태로 얻어질 때까지 용액을 계속해서 가열하여 교반하였다. 분말 형태로 얻어진 뒤, 14 g의 요소(Urea, (NH₂)₂CO)를 넣고 용액을 사용하여 섞어주었다. 이 혼합물을 세라믹 제질의 뚜껑이 닫힌 도가니 안에 넣고, 분당 5 ℃의 승온 속도로 550 ℃의 박스형 전기로에서 3 시간 동안 열분해를 하였다. 열분해 후 얻어지는 분말은 증류수로 3 번의 세척 후 여과한 후, 80 ℃의 오븐에서 12 시간 동안 건조하였다. 인 도량 농도는 제2 원산암모늄의 양을 통해 조절하여 PgCN 물질을 합성하였다.

[0061] **실시예 2: 인 도량 조절된 탄소결합물 (PgCN, PgCN)의 합성**

[0062] 6 g의 멜라민(Melamine) 분말을 1000 mL 증류수가 담긴 비커에 넣고, 100 ℃로 설정된 핫플레이트 위에서 멜라민이 완전히 녹을 때까지 60분 동안 교반하였다. 그 다음 60% 농도의 질산 용액을 5 mL 넣고, 용액이 모두 증발하여 건조한 분말 형태로 얻어질 때까지 용액을 계속해서 가열하여 교반하였다. 분말 형태로 얻어진 뒤, 14 g의 요소(Urea, (NH₂)₂CO)와 6 g의 제2 원산암모늄((NH₄)₂HP₂O₇)을 추가로 넣고 용액을 사용하여 섞어주었다. 이 혼합물을 세라믹 제질의 뚜껑이 닫힌 도가니 안에 넣고, 분당 5 ℃의 승온 속도로 550 ℃의 박스형 전기로에서 3 시간 동안 열분해 하였다. 열분해 후 얻어지는 분말은 증류수로 3 번의 세척 후 여과한 후, 80 ℃의 오븐에서 12 시간 동안 건조하였다. 인 도량 농도는 제2 원산암모늄의 양을 통해 조절하여 PgCN 물질을 합성하였다.

[0063] **실시예 3 내지 4: PgCN/C 또는 gCN/C에 코팅된 분리막의 제조**

[0064] (1) 먼저 슬러리 전체 중량을 기준으로 60 중량%의 PgCN(또는 gCN)과 30 중량%의 탄소 혼합물, 10 중량%의 PVDF 매진제 조성으로 이루어진 슬러리를 30 분 동안 밀링을 통해 제조하였다. 여기서 30 중량%의 탄소 혼합물의 세부 조성은 제조된 전체 슬러리 100 중량부 기준으로 20 중량부의 Super-P, 5 중량부의 VGCF(Vapor Grown Carbon Fiber), 5 중량부의 rGO(Reduced Graphene Oxide)로 하였다.

[0065] (2) 이렇게 제조된 슬러리를 다투 분레이드 방식으로 분리막(폴리프로필렌) 한쪽 표면 위에 캐스팅한 뒤, 40 ~ 50 ℃의 오븐에서 12 시간 동안 건조하여 PgCN/C 코팅된 분리막을 제작하였다. gCN/C 코팅된 분리막 제작 방법은 PgCN/C 대신 gCN/C 물질을 사용하고, 나머지 절차는 동일하다.

도면1 중량비

(a)

Samples	Melamine/Urea ratio				
	9 / 1	7 / 3	5 / 5	3 / 7	1 / 9
BET S.A. (m ² /g)	9	15	19	52	53

(b)

(c)

도면2 함량비

Name	M/U (g)	Ammonium salt (g)	Atomic P percent
PgCN7	6/14	2	7.2
PgCN14	6/14	4.5	14
PgCN22	6/14	6	22.8
PgCN29	6/14	7.5	29
PgCN36	6/14	9	36

그림 27. 이차전지 큐레이션 대상 항목 예시

○ 첨단바이오

표 15. 첨단바이오 분야 과제 개요

구분	상세 내용
기술개요	• 본 기술은 생명현상을 공학적으로 설계·제작하는 합성생물학과 환자 맞춤형 치료를 실현하는 유전자·세포치료, 그리고 데이터 기반의 디지털 헬스케어 기술을 포함하며, 특히 기존 약물로 치료가 불가능한 난치병 대응을 위한 차세대 바이오 의약품 개발에 집중함
필요성	• (바이오-디지털 컨버전스 대응) 알파폴드 등 AI 기반 데이터 중심으로의 R&D 패러다임 전환에 맞서, 글로벌 빅테크의 지식 독점에 대응할 고순도 공공 데이터셋 구축이 시급함 • (상용화 병목 및 불확실성 해소) TRL 연계과정중 발생하는 생체 반응 및 공정 변수의 불일치를 극복하기 위해, 비정형 임상·실험 데이터를 자산화하여 최적 경로를 예측해야 함 • (데이터 상호운용성 및 주권 확보) 국내 바이오 데이터의 글로벌 경쟁력 제고를 위해 OMOP-CDM 등 국제 표준 코드 체계와 호환되는 IP 데이터 큐레이션 체계 수립이 필수적임

범위 및 목적	• (최종 목표) 첨단바이오 4대 핵심 분야의 특허 빅데이터를 자산화하여, TRL 4단계 이상의 시작품 또는 치료제의 효능 예측 정밀도를 향상시키고 제품 또는 기술개선을 20% 이상 제고할 수 있는 지능형 R&D 솔루션 개발 • (지원 범위) TRL 4단계(실험실 수준의 시운전 및 시작품 제작) 이상의 상용화 연계 기술 개발 및 실증 데이터셋 구축 • (대상 분야) 합성생물학, 유전자·세포치료, 감염병 백신·치료, 디지털 헬스데이터 분석·활용
세부 기술 요구사항	① 합성생물학 (바이오파운드리, 유전자 회로 설계 등) • 데이터 추출 요구사항 - DNA/RNA 염기서열 설계 데이터 및 단백질 발현 효율(IC₅₀, EC₅₀ 등) 수치 큐레이션 - 대사경로 최적화를 위한 효소 활성 지표 및 미생물 균주별 배양 조건(pH, 온도, 용존산소 등) 데이터 구조화 • 실증 내용 - 특허 기반 설계 데이터를 적용한 고부가가치 화합물 생산용 세포공장 시작품의 생산 수율이 기존 대비 15% 이상 향상됨을 실증함 - R&D솔루션이 제안한 유전자 회로의 작동 안정성이 연속 배양 공정에서 유지됨을 데이터로 증명함 ② 유전자·세포치료 (CAR-T, 줄기세포, 엑소좀 치료제 등) • 데이터 추출 요구사항 - 세포주 특성 데이터(CD 마커 등) 및 유전자 전달체(Viral Vector 등)의 형질전환 효율 데이터 큐레이션 - 파라클라인 효과 인자(VEGF, IGF-1 등) 분비량과 치료 효능 지표 간의 상관관계 데이터 추출 • 실증 내용 - 세포치료제 시작품의 순도(Purity) 및 역가(Potency) 시험 결과가 특허 내 최상위 기술 규격 대비 동등 이상임을 입증함 - 특허 기반 공정 최적화 데이터 적용 시, 치료제의 세포 생존율(Viability)이 90% 이상 확보됨을 실측 데이터로 검증함 ③ 감염병 백신·치료 (mRNA 플랫폼, 항바이러스제 등) • 데이터 추출 요구사항 - 항원 설계 서열 및 유효입자 캡슐화 효율, 입자 크기 안정성 데이터 큐레이션 - 전임상/임상 단계의 중화항체 역가(GMT) 및 면역원성 데이터의 비정형 수치 구조화 • 실증 내용 - 시작품 단계의 백신 후보물질이 동물 모델 시험에서 특허 데이터 기반 예측치와 일치하는 항체 형성 유도능을 보유함을 실증 - mRNA 시작품의 가속 가혹 시험을 통해 특허 기반 안정성 데이터와의 부합성을 증명 ④ 디지털 헬스데이터 분석·활용 (AI 진단, 디지털 치료제 등) • 데이터 추출 요구사항 - 질병 진단 알고리즘의 민감도, 특이도 및 AUC 지표 큐레이션 - 임상 용어(SNOMED-CT), 검사항목(LOINC) 등 국제 표준 코드와 매핑된 특허 내 증상·진단 데이터셋 구축 • 실증 내용 - 구축된 데이터셋으로 학습된 AI 진단 모델이 실제 임상 환경 모사 데이터에서 기존 알고리즘 대비 정확도가 10% 이상 향상됨을 확인 - OMOP-CDM 표준 구조를 적용한 특허 데이터셋이 병원 정보시스템과의 상호운용성을 확보했음을 테스트 베드 실증을 통해 입증함

발명의 내용

해결하려는 과제

본 발명은, 상기 사제를 감압하여 이루어진 것으로, 패턴 형성 시의 해상성을 향상하고, 또한, 라인 엣지 리프 니스(LEER)가 저감된 패턴을 얻을 수 있는 화학 증착 네거티브용 레지스트 조성물, 이것을 도포한 포토마스크를 행크, 및 레지스트 패턴 형성 방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

본 발명자들은, 상기 목적을 달성하기 위해서 예의 검토를 거듭한 결과, 질소를 함유하는 카르복실산의 술포늄염을 레지스트 조성물에 도입한 경우, LEER이 작은 패턴이 얻어지는 것을 발견하여, 본 발명을 완성하기에 이르렀다.

따라서, 상기 과제를 해결하기 위해서, 본 발명은,

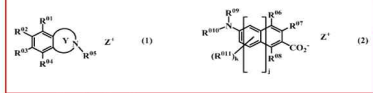
(A) 하기 식 (1) 또는 식 (2) 중 어느 구조로 표시되는 **소염염**

성분

(B) 하기 식 (3)으로 표시되는 반복 단위를 갖는 수지(D-1),

(C) 가교제

를 함유하는 코에너지전달 노광용 화학 증착 네거티브용 레지스트 조성물을 제공한다.



구조식

[식에서, R^{1a}, R^{1b}, R^{1c}, R^{1d}는 각각 독립적으로 수소 원자, -L-CO₂⁻, 혹은 해테로 원자로 치환되어 있어도 좋고, 해테로 원자가 개재하여도 좋은 탄소수 1~20의 직쇄상, 또는 탄소수 3~20의 분기상 또는 환상의 1가 탄화수소기를 나타내고, R^{2a}와 R^{2b}, R^{2c}와 R^{2d}, R^{2e}와 R^{2f} 중 어느 조합에 있어서, 상호 결합하여 이들이 결합하는 탄소 원자와 함께 고리를 형성하여도 좋다. L은 단일 결합을 나타내거나 혹은 해테로 원자로 치환되어 있어도 좋고, 해테로 원자가 개재하여도 좋은 탄소수 1~20의 직쇄상, 또는 탄소수 3~20의 분기상 또는 환상의 2가 탄화수소기를 나타낸다. R³는 수소 원자 또는 해테로 원자로 치환되어 있어도 좋고, 해테로 원자가 개재하여도 좋은 탄소수 1~20의 직쇄상, 또는 탄소수 3~20의 분기상 또는 환상의 1가 탄화수소기를 나타낸다. R^{4a}, R^{4b}, R^{4c}, R^{4d}, R^{4e} 및 R^{4f}은 각각 독립적으로 수소 원자, 혹은 해테로 원자로 치환되어 있어도 좋고, 해테로 원자가 개재하여도 좋은 탄소수 1~20의 직쇄상, 또는 탄소수 3~20의 분기상 또는 환상의 1가 탄화수소기를 나타내고, R^{5a}와 R^{5b}, R^{5c}와 R^{5d}, R^{5e}와 R^{5f} 중 어느 조합에 있어서, 상호 결합하여 이들이 결합하는 탄소 원자와 이들 사이의 탄소 원자와 함께 고리를 형성하여도 좋다. j는 0 또는 1이며, j=0인 경우는 0≤k≤1, j=1인 경우는 0≤k≤3이다. Z⁺는 하기 식 (a)로 표시되는 술포늄 양이온 또는 식 (b)로 표시되는 요오드늄 양이온을 나타낸다.

발명을 실시하기 위한 구체적인 내용

본 발명자들은, 상기 목적을 달성하기 위해서 예의 검토를 거듭한 결과, 질소를 함유하는 카르복실산의 술포늄염을 레지스트 조성물에 도입한 경우, LEER이 작은 패턴을 얻을 수 있는 것을 발견하여, 본 발명을 완성하기에 이르렀다.

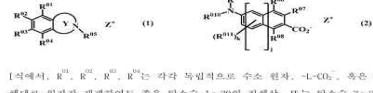
이하, 본 발명에 관해서 상세히 기술한다.

한편, 이하의 설명 중, 화학식으로 표시되는 구조에 따라서는 비대칭 탄소 존재하고, 엔안티오머(enantiomer)나 다이아스테레오머(diastereomer)가 존재할 수 있는 것이 있지만, 그 경우는 하나의 식으로 이들 이성체를 대표하여 나타낸다. 이들 이성체는 단독으로 이용하여도 좋고, 혼합물로서 이용하여도 좋다.

본 발명은,

(A) 하기 식 (1) 또는 식 (2) 중 어느 구조로 표시되는 소염염,

(B) 하기 식 (3)을 포함하는, 상기 작용에 의해 알칼리 불용성이 되는 수지를 함유하는 코에너지전달 노광용 화학 증착 네거티브용 레지스트 조성물을 제공한다.

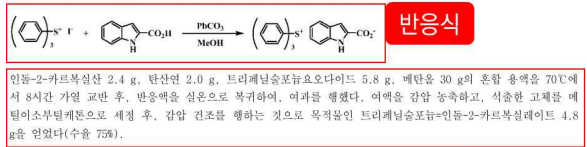


[식에서, R^{1a}, R^{1b}, R^{1c}, R^{1d}는 각각 독립적으로 수소 원자, -L-CO₂⁻, 혹은 해테로 원자로 치환되어 있어도 좋고, 해테로 원자가 개재하여도 좋은 탄소수 1~20의 직쇄상, 또는 탄소수 3~20의 분기상 또는 환상의 1가 탄화수소기를 나타내고, R^{2a}와 R^{2b}, R^{2c}와 R^{2d}, R^{2e}와 R^{2f} 중 어느 조합에 있어서, 상호 결합하여 이들이 결합하는 탄소 원자와 함께 고리를 형성하여도 좋다. L은 단일 결합을 나타내거나, 혹은 해테로 원자로 치환되어 있어도 좋고, 해테로 원자가 개재하여도 좋은 탄소수 1~20의 직쇄상, 또는 탄소수 3~20의 분기상 또는 환상의 2가 탄화수소기를 나타낸다. R³는 수소 원자 또는 해테로 원자로 치환되어 있어도 좋고, 해테로 원자가 개재하여도 좋은 탄소수 1~20의 직쇄상, 또는 탄소수 3~20의 분기상 또는 환상의 1가 탄화수소기를 나타낸다. R^{4a}, R^{4b}, R^{4c}, R^{4d}, R^{4e} 및 R^{4f}은 각각 독립적으로 수소 원자, 혹은 해테로 원자로 치환되어 있어도 좋고, 해테로 원자가 개재하여도 좋은 탄소수 1~20의 직쇄상, 또는 탄소수 3~20의 분기상 또는 환상의 1가 탄화수소기를 나타내고, R^{5a}와 R^{5b}, R^{5c}와 R^{5d}, R^{5e}와 R^{5f} 중 어느 조합에 있어서, 상호 결합하여 이들이 결합하는 탄소 원자와 이들 사이의 탄소 원자와 함께 고리를 형성하여도 좋다. j는 0 또는 1이고, j=0인 경우는 0≤k≤1, j=1인 경우는 0≤k≤3이다.

실시예

이하, 실시예 및 비교예를 기재하여 본 발명을 구체적으로 설명하지만, 본 발명은 하기의 실시예로 제한되는 것은 아니다. 한편, 하기 예 중, Me는 메틸기를 나타낸다. 또한, 공중합 조성비는 분비이며, 중량 평균 분자량(Mw)은 겔 침투 크로마토그래피(GPC)에 의한 폴리스티렌 환산 중량 평균 분자량을 나타낸다.

[합성예 1-1] 트리페닐술포늄=인돌-2-카르복실레이트(염-1)의 합성



얻어진 목적물의 스펙트럼 데이터를 하기에 나타낸다.

반응 조건

적외 흡수 스펙트럼(IR(D-ATR): cm⁻¹)

3389, 3135, 3083, 3022, 1574, 1523, 1474, 1445, 1415, 1382, 1371, 1340, 1320, 1230, 995, 844, 826, 816, 765, 755, 748, 699, 683 cm⁻¹.

비행 시간강 질량 분석(TOFMS: MALDI)

POSITIVE M⁺ 263 ((C₁₆H₉)₃S⁺ 상당)

NEGATIVE M⁻ 160 (C₈H₇N-CO₂⁻ 상당)

[합성예 2-3] 폴리머 2-7, 9-16의 합성

각 단량체의 종류, 도입비(몰비)를 바꾼 것 이외에는 폴리머 1, 8과 동일한 수순에 의해, 표 1에 나타낸 폴리머 2-7, 9-16을 합성했다.

표 1에 폴리머 1-16의 단량체의 종류와 도입비를 통합하여 나타낸다. 또한, 폴리머에 도입한 반복 단위의 구조를 표 2-5에 나타낸다.

표 1

조성비

	단량1	도입비 (mol%)	단량2	도입비 (mol%)	단량3	도입비 (mol%)	단량4	도입비 (mol%)
폴리머1	A-1	80.0	B-1	10.0	B-5	10.0		
폴리머2	A-1	80.0	B-2	8.0	B-4	12.0		
폴리머3	A-1	60.0	B-2	10.0	C-1	30.0		
폴리머4	A-1	70.0	B-2	7.0	C-2	23.0		
폴리머5	A-1	70.0	B-2	10.0	C-3	20.0		
폴리머6	A-1	70.0	B-2	10.0	C-4	20.0		
폴리머7	A-1	55.0	B-3	10.0	C-1	35.0		
폴리머8	A-1	66.0	B-2	9.0	C-1	21.5	P-1	3.5
폴리머9	A-1	63.0	B-2	5.0	C-1	24.0	P-1	8.0
폴리머10	A-1	67.0	B-2	10.0	C-1	18.5	P-2	4.5
폴리머11	A-1	67.0	B-2	9.3	C-1	20.0	P-3	3.7
폴리머12	A-1	67.3	B-2	10.0	C-1	17.5	P-4	5.2
폴리머13	A-1	64.1	B-2	9.5	C-1	22.0	P-5	4.4
폴리머14	A-1	64.0	B-2	10.0	C-1	22.8	P-6	3.2
폴리머15	A-1	62.0	B-3	10.0	C-1	24.3	P-1	3.7
폴리머16	A-2	60.5	B-4	10.0	C-1	24.4	P-2	5.1

그림 28. 첨단바이오 큐레이션 대상 항목 예시

○ ICT표준-차세대 통신

표 16. ICT표준-차세대 통신 분야 과제 개요

구분	상세 내용
기술개요	본 기술은 초대역·초저지연·초연결 실현을 위한 6G 이동통신과 지상망의 한계를 극복하는 5G-6G 위성통신(NTN), 특정 밴드 종속성을 탈피하는 Open-RAN(개방형 무선접속망), 그리고 이를 저전력으로 구동하기 위한 고효율 통신 부품 기술 등을 포함
필요성	(6G 기술 주도권 및 SEP 확보) 5G를 압도하는 전송 속도와 커버리지 확보를 위한 글로벌 표준 경쟁 속에서, 국제 표준 문서와 특허에 산재한 기술 정보를 자산화하여 국내 기업의 핵심 IP 확보 지원 (지상-위성 통합 및 패러다임 전환) NTN 통합 및 Open-RAN 국산화를 위해 TRL 3 수준의 알고리즘이 실제 고주파-위성 환경(TRL 4)에서 겪는 신호 감쇄 및 도플러 효과 등을 특허 데이터를 통해 사전 예측하고 해결함.
범위 및 목적	(최종 목표) 차세대 통신 4대 핵심 분야의 특허 빅데이터를 자산화하여, 목표제품 및 기술의 통신 전송 효율과 전력 성능을 예측하고 기술 구현 성공률을 20% 이상 제고할 수 있는 지능형 R&D 솔루션 구축 (지원 범위) TRL 4단계(통신 환경 시험 및 통신 모듈 시작품 성능 검증) 이상의 상용화 연계 기술 개발 및 실증 데이터셋 구축 (대상 분야) 5G-6G 위성통신, 6G(또는 5G) 이동통신, Open-RAN, 고효율 5G-6G 통신부품

세부 기술 요구사항	① 5G-6G 위성통신 및 6G 이동통신 (NTN, sub-THz, 빔포밍 등) • 데이터 추출 요구사항 - 저궤도 위성망의 링크 연결 설계, 궤도 파라미터 및 도플러 시프트 보상 알고리즘 데이터 큐레이션 - sub-THz 대역의 채널 모델링 변수 및 대규모 다중 입출력 빔포밍 가중치 수치 데이터 구조화 • 실증 내용 - 특히 기반 예측 모델을 통해 산출된 위성-지상 간 핸드버 성공률과 실제 단계에서의 환경 시험(시뮬레이터 연동 등) 결과 간의 오차율 10% 이내 검증 - 6G 시작품의 데이터 전송률 및 지연시간 지표가 특히 내 최상위 기술 규격 대비 동등 이상임을 실측 데이터로 입증함
	② Open-RAN (개방형 인터페이스, 가상화 RAN 등) • 데이터 추출 요구사항 - O-RAN 얼라이언스 표준 인터페이스(E2, F1, 7-2x split 등) 규격 및 상호운용성 테스트 시나리오 데이터 큐레이션 - 지능형 무선접속망 컨트롤러(RIC)의 자원 할당 최적화 로직 및 xApp/rApp 연동 파라미터 구조화 • 실증 내용 - 이기종 벤더 간 장비 연동 시 인터페이스 적합성이 특히 데이터 기반 표준 가이드라인 준수를 기준 시험환경 기준 테스트 베드에서 실증함 - 가상화된 RAN 소프트웨어의 처리 오버헤드가 특히 기반 예측치 대비 5% 이내의 정밀도를 가짐을 실측 데이터로 증명함
	③ 고효율 5G-6G 통신부품 (RFIC, 전력증폭기, 안테나 소자 등) • 데이터 추출 요구사항 - GaN-on-SiC 등 화합물 반도체 기반 전력증폭기(PA)의 전력부가효율(PAE) 및 선형성 지표 데이터 큐레이션 - 안테나 소자의 이득(Gain), 방사 패턴 및 초고주파 패키징용 저손실 소재의 유전율 데이터 구조화 • 실증 내용 - 특히 기반 최적 설계 데이터를 적용한 RF 부품(대조군 제품 기준)의 전력 소모량이 기존 대비 15% 이상 저감됨을 기준 환경 시험을 통해 입증함 - 고주파 구동 시 발생하는 열 변형과 신호 간섭 수치가 특히 데이터 기반 시뮬레이션 모델 범위 내에 있음을 검증함

○ ICT표준-AI(인공지능)

표 17. ICT표준-차세대 통신 분야 과제 개요

구분	상세 내용
기술개요	• 본 기술은 대규모 데이터를 학습하고 추론하기 위한 효율적 AI 인프라와 인간 수준의 판단력을 갖춘 첨단 모델링, 그리고 산업 현장에서 신뢰할 수 있는 안전·신뢰 AI 기술을 의미하며, 소프트웨어 알고리즘과 하드웨어 가속기의 최적 결합을 지향함
필요성	• (AI 주권 확보 및 R&D 환경 조성) LLM 및 생성형 AI 확산에 대응하여 글로벌 빅테크의 핵심 알고리즘과 인프라 설계 특허를 자산화하고, 초격차 기술 확보를 위한 데이터 기반 R&D 환경을 조성함 • (현장 적용 난제 및 병목 해소) 알고리즘 기술이 적용될 때 발생하는 추론 속도 및 호환성 문제를 해결하기 위해, 특히 내 최적화 파라미터를 활용한 성능 예측 체계를 구축함 • (글로벌 규제 대응 및 신뢰성 확보) EU AI Act 등 국제적 AI 안전 규제에 선제적으로 대응하기 위해, 표준화된 안전성 지표를 반영한 데이터 큐레이션 및 기술 실증 체계 수립이 필수적임

범위 및 목적	• (최종 목표) 인공지능 분야의 특허 빅데이터를 자산화하여, AI의 성능(정확도, 속도 등) 예측 정밀도를 높이고 실무 적용 가능성을 20% 이상 향상할 수 있는 지능형 솔루션 개발 및 기술개선 • (지원 범위) TRL 4단계(특정 산업 데이터 기반의 AI 모델 구현 및 시작품 성능 검증) 이상의 상용화 연계 기술 개발 및 실증 데이터셋 구축 • (대상 분야) 첨단 AI 모델링·의사결정, 산업 활용·혁신 AI
세부 기술 요구사항	① 첨단 AI 모델링·의사결정 (멀티모달, 에이전트 AI, 자율 의사결정 등) • 데이터 추출 요구사항 - 멀티모달 데이터 통합을 위한 어텐션(Attention) 가중치 분포 및 토큰 최적화 알고리즘 데이터 큐레이션 - 자율 에이전트의 의사결정 트리 및 강화학습 보상 함수(Reward Function) 설계 변수 추출 • 실증 내용 - 특허 기반의 하이퍼파라미터 큐레이션 데이터를 적용한 시작품 모델의 정확도(F1-score 등)가 선행 기술 대비 동등 이상임을 입증함 - 복합 상황에서의 의사결정 지연 시간(Latency)이 특정 단계 테스트 베드에서 가이드라인 기준의 만족을 실측 데이터로 확인함 ② 산업 활용·혁신 AI (제조, 금융, 의료 특화 AI 등) • 데이터 추출 요구사항 - 산업별 특화 데이터 전처리 로직 및 도메인 지식 그래프(Knowledge Graph) 구성 변수 큐레이션 - 실시간 OT(Operational Technology) 데이터와 AI 모델 간의 연동 인터페이스 표준 및 프로토콜 데이터셋 구축 • 실증 내용 - 제조 또는 금융 등 특정 산업 도메인 시작품에 특허 기반의 혁신 알고리즘을 적용하여, 기존 수작업 대비 업무 효율이 20% 이상 개선됨을 실증함 - 가상 시뮬레이션 환경(Digital Twin)에서 학습된 AI 모델이 실제 현장 데이터 적용 시 성능 유지력을 보유함을 증명

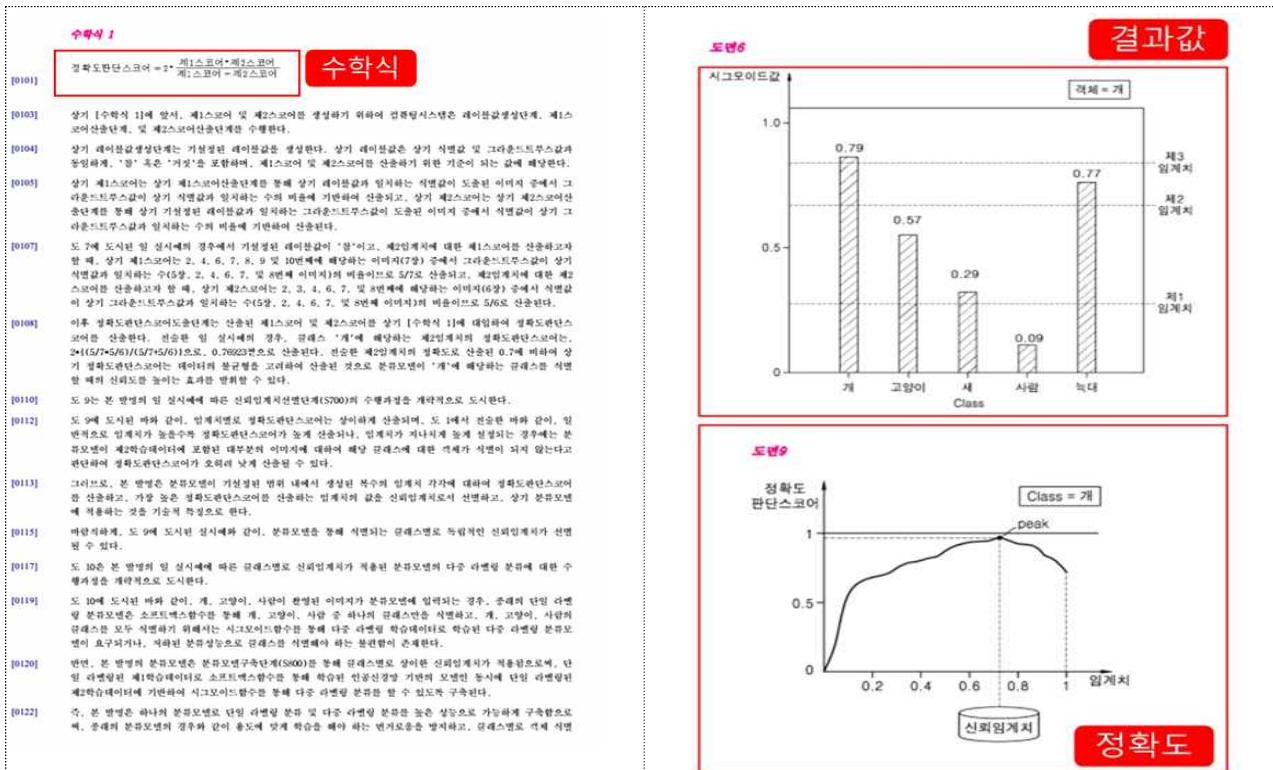


그림 29. 인공지능 큐레이션 대상 항목 예시

○ ICT표준-첨단로봇·제조

표 18. ICT표준-첨단로봇·제조 분야 과제 개요

구분	상세 내용
기술개요	본 기술은 제조 현장의 자동화를 넘어 인간과 협업하고 스스로 환경을 인지하여 이동·조작하는 지능형 로봇과, 가상 세계와 실제 현장을 동기화하는 디지털 트윈 기반 가상 제조 기술을 포괄함
필요성	- 제조 패러다임 변화 및 로봇 지능화 – 노동력 부족 해소를 위한 자율 조작 및 협업 기술이 핵심인 가운데, 글로벌 선도 기업의 제어 알고리즘과 부품 설계 특허를 자산화하여 국내 기술 격차를 단기간에 해소해야 함 - Sim2Real 간극 및 실증 장벽 해소 – 시뮬레이션과 실제 환경 간의 제어 오차를 극복하기 위해, 특허 내 물리 특성치와 파라미터를 데이터화하여 시작품의 성능을 사전 예측하는 체계 구축이 시급함 - 국제 표준 기반 안전 및 상용화 확보 – 인간-로봇 상호작용(HRI)의 안전성 확보를 위해 ISO 10218 등 국제 표준을 반영한 데이터 큐레이션 및 기술 실증을 상용화의 필수 전제로 추진함
범위 및 목적	(최종 목표) 첨단로봇·제조 지능화 분야의 특허 빅데이터를 자산화하여, TRL 4단계 시제품의 작동 정밀도와 자율성을 향상시키고, 제조공정 효율을 20% 이상 제고할 수 있는 지능형 솔루션 개발 및 기술개선 (지원 범위) TRL 4단계(특정 작업 환경에서의 로봇 시작품 구동 및 가상 제조 연동 검증) 이상의 상용화 연계 기술 개발 및 실증 데이터셋 구축 (대상 분야) 로봇 정밀제어·구동 부품·SW, 로봇 자율이동 및 고난도 자율조작, 인간-로봇 상호작용
세부 기술 요구사항	① 로봇 정밀제어·구동 부품·SW (감속기, 서보모터, 제어 알고리즘 등) - 데이터 추출 요구사항 - 고정밀 감속기의 치형 설계 변수, 토크 밀도 및 백래시(Backlash) 억제 기술 관련 수치 데이터 큐레이션 - 모터 제어를 위한 PID/MPC 알고리즘의 이득(Gain) 파라미터 및 전류 제어 루프의 응답성 지표 구조화 - 실증 내용 - 특허 기반의 최적 설계 데이터를 적용한 구동 부품 시작품의 회전 정밀도 및 에너지 효율이 기존 제품 대비 10% 이상 향상됨을 실증함 - 고속 구동 시 발생하는 진동 및 소음 수치가 특허 데이터 기반 예측 모델 범위 내에 있음을 시운전을 통해 검증함 ② 로봇 자율이동 및 고난도 자율조작 (SLAM, 물체 인식, 그리퍼 등) - 데이터 추출 요구사항 - 자율주행을 위한 SLAM(위치 추정 및 지도 작성) 알고리즘의 오차 보정 계수 및 경로 최적화 변수 데이터 큐레이션 - 다양한 형태의 물체 파지(Grasping)를 위한 그리퍼의 압력 제어 수치 및 촉각 센서 데이터 처리 로직 추출 - 실증 내용 - 특허 기반의 환경 인지 알고리즘을 탑재한 이동 로봇 시작품의 위치 추정 오차가 실내 환경 기준 5cm 이내임을 실증함 - 비정형 물체에 대한 자율 조작 성공률이 특정 테스트 베드에서 90% 이상 달성됨을 데이터로 입증함

③ 인간·로봇 상호작용 및 가상 제조 (HRI, 디지털 트윈, 협동 로봇 등) • 데이터 추출 요구사항 - 인간의 의도 파악을 위한 음성/제스처 인식 알고리즘의 인식을 지표 및 충돌 방지 안전 거리 설정 데이터 큐레이션 - 가상 제조 환경 구현을 위한 로봇-설비 간 데이터 인터페이스 표준 및 시뮬레이션-실제 간 데이터 동기화(Sync) 변수 구조화 • 실증 내용 - ISO 10218(로봇 안전) 표준을 반영한 안전 제어 로직이 적용된 시작품의 충돌 감지 및 정지 반응 시간이 가이드라인 기준을 준수함을 실측함 - 가상 제조 모델과 실제 제조 라인 간의 물리적 거동 일치율이 95% 이상임을 디지털 트윈 연동 테스트를 통해 증명함	
---	--

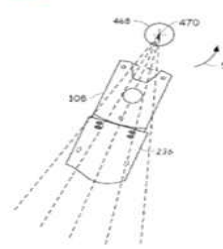
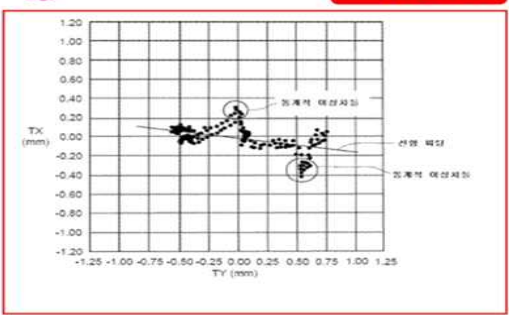
[3698] 일 실시 예에서, 재료 투입 구동부(250)는 상기 재료 공급부(230), 무게 측정부(280) 및 재료 토출부(240)가 동작하도록 구동력을 제공할 수 있다. 예를 들어, 재료 투입 구동부(250)는 재료 공급부(230), 무게 측정부(280), 및 재료 토출부(240)의 작동에 구동력을 제공할 수 있다. 재료 투입 구동부(250)는 제어부(100)와 연동되어 각 구성 요소가 최적의 시점에 작동하도록 한다. 이 구동부는 전기 모터와 같은 정밀 제어 장치를 포함하여 재료의 이동과 투입 과정이 정확하게 이루어질 수 있도록 설계될 수 있다. [3699] 본 발명의 재료 투입부(300)은 이러한 구성 요소들이 유기적으로 결합되어 작동하며, 온도 제어 과정에서 재료 투입의 정확성과 효율성을 크게 향상시킬 수 있다. 이를 통해 사용자는 대량 용기를 조리할 때 투입 과정에서 발생할 수 있는 오류를 줄이고, 전체 조리 품질을 보장할 수 있다. [3700] 한편, 일 실시 예에서, 제어부(100)는 온도의 실시간 모니터링을 상기 재료 투입부(300)를 통해 측정한 재료들의 투입량 및 투입 시기를 제어할 수 있다. [3701] 또한, 복수의 재료들의 투입량을 조절하는 동작들을 설명하면 다음과 같다. [3702] 일 실시 예에서, 제어부(100)는 동작 11에서, 미리 설정된 투입량을 확인할 수 있다. 예를 들어, 제어부(100)는 사용자가 미리 입력한 레시피 데이터를 참조하여 각 재료의 투입량을 확인할 수 있다. 이 과정에서 제어부(100)는 재료별로 설정된 목표 투입량과 투입 시기를 기반으로 작업을 계획할 수 있다. 설정된 투입량은 제어부(100)에 저장된 데이터베이스 또는 외부 입력 장치(500)를 통해 입력된 레시피 데이터를 포함할 수 있다. [3703] 그리고 제어부(100)는 동작 12에서, 상기 무게 측정부를 통해 측정된 실제 투입량과 상기 미리 설정된 투입량을 비교할 수 있다. 예를 들어, 제어부(100)는 재료 투입부(200)에 포함된 무게 측정부(280)를 통해 측정된 실제 투입량을 확인할 수 있다. 측정된 값은 목표가 무게 측정부(280)를 통과하는 동안 실시간으로 감지되며, 제어부(100)는 이를 설정된 목표 투입량과 비교할 수 있다. 이 비교 과정에서 설정된 허용 오차 범위를 초과하거나 부족한 경우, 적절한 보정 조치를 수행할 수 있다. [3704] 동작 13에서, 비교 결과, 상기 실제 투입량이 상기 미리 설정된 투입량보다 작은 경우, 제어부(100)는 상기 복수의 재료들을 추가로 공급할 수 있다. 예를 들어, 제어부(100)가 추가 공급 명령을 재료 공급부(230)와 재료 투입 구동부(250)에 전달하여 부족한 재료를 보충할 수 있다. 추가 재료 공급은 설정된 목표 투입량에 도달할 때까지 반복적으로 수행될 수 있으며, 이 과정에서 무게 측정부(280)가 지속적으로 실시간 값을 측정하여 제어부(100)에 데이터를 제공할 수 있다. [3705] 여기서 동작 14에서, 제어부(100)는, 상기 재료 투입부(300)를 통해 미리 설정된 투입 시기에 따라 상기 복수의 재료들을 상기 조리 용기에 투입할 수 있다. 예를 들어, 제어부(100)는 재료 토출부(240)를 활성화하여 재료는 조리 용기로 이동시키는 동작을 제어할 수 있다. 이 과정은 재료의 혼합 상태를 최소화하고, 조리 용기로 투입된 순서를 유지할 수 있도록 설계될 수 있다. [3706] 한편, 복수의 재료들의 투입 시기를 조절하는 동작들을 설명하면 다음과 같다. [3707] 일 실시 예에서, 동작 21에서, 제어부(100)는 상기 온도의 조리 단계에서 요구되는 목표 온도, 상기 온도의 현재 온도 및 온도 변화 계수에 기반하여 상기 복수의 재료들의 투입 시기를 조절할 수 있다. 예를 들어, 제어부(100)는, 상기 식 1에 따라 상기 복수의 재료들의 조절된 투입 시기를 연산할 수 있다. [3708] [식 1] $t_{adjust, i} = t_{base, i} + \Delta t_i, \quad \Delta t_i = \frac{(T_{target} - T_{current})}{k_T}$ [3709] 동작제어 수식	도면 5  이동 범위 
---	---

그림 30. 첨단로봇제조 큐레이션 대상 항목 예시

□ 비정형 데이터셋 구축의 유형별 핵심 알고리즘(예)

- 화학 구조식, 도면, 서열 등 각 유형에 특화된 인식·추출 알고리즘을 분리 적용

표 19. 분야별 비정형 데이터셋

유형	분야	핵심 알고리즘 설계	R&D 활용(예)
화학 구조식	바이오·이차전지	CNN 인코더+Transformer 디코더로 이미지 → SMILES 변환 → Molecular GNN으로 분자 지문 벡터 생성	구조-활성 관계(SAR) 데이터셋 → 신규 후보물질 역설계
도면·그래프	전 분야	허프 변환 축 탐지+ 도메인 특화 OCR + 색상 분리·곡선 추적으로 데이터 포인트 수치 역변환	특허 실험 결과 대규모 추출 → AI 성능 예측 모델 학습
수식·수치 범위	반도체·ICT	im2latex 수식 인식 + 의존 구문 분석으로 파라미터-성능 쌍 구조화 JSON 추출	공정 파라미터 최적화 조건 도출
서열 정보	첨단바이오	WIPO ST.26/ST.25 파서 + NCBI/UniProt ID 매핑 자동화 + 서열-기능 관계 NER	유전자·단백질 기능 DB → 치료제 효능 예측

□ 국가전략기술 분야별 기술 난제 해결 알고리즘(예)

- 기술 특성과 데이터 구조가 다르므로, 각 분야의 핵심 난제에 맞춰 알고리즘 구성과 기대 효과를 달리 설계

표 20. 분야별 기술 난제 해결 알고리즘

분야	난제(예)	알고리즘 구성	기대 효과
반도체·디스플레이	공정 레시피 추출 및 수율 예측	공정 흐름도→그래프 변환(비전-언어 모델) + BERT 공정 시퀀스 임베딩 + 멀티태스크 회귀 역설계	목표 성능(HBM 대역폭 등) 달성 공정 조건 자동 추천
첨단바이오	치료 효능 예측 정확도 향상	화합물-활성DB 자동 구축(SMILES+SAR) + 전이 학습 효능 예측 모델 파인튜닝 + 멀티오믹스 통합	후보물질 효능 사전 예측 → 임상 실패율 감소
이차전지	소재 조성 최적화	조성-성능 데이터 표준화(도메인 적응) + GCN 기반 결정 구조 예측 + 미탐색 고성능 조성 구간 도출	에너지밀도·수명 동시 최적화 소재 조성 제안
표준ICT(통신·로봇)	표준 특허 적합성 자동 분석	표준문서-특허 청구항 의미 유사도 임베딩+SEP 후보 자동 식별 + 표준 진화 포트폴리오 모니터링	표준기여 기회(White Space) 발굴 → 선제적 SEP 창출

제4절 추진전략

□ [추진전략1] 전략적 IP 데이터 자산화 체계 확립

- (추진목표) 비정형 지식재산(IP) 데이터를 AI가 즉시 연산·추론 가능한 고순도 디지털 연구 자산으로 정밀 구조화

[세부전략1-1] 도메인 특화 딥 큐레이션 수행

- 비정형 정보의 데이터화 : 특허 명세서 내에 매몰되어 있던 마쿠쉬 구조, 복잡한 실험 테이블, 시계열 수치 데이터 등을 AI가 학습이 가능한 표준 구조화 데이터(JSON/XML 등)로 정밀 추출
- 핵심 파라미터 타겟팅 : 단순 서지 정보 검색에서 벗어나 연구 현장의 실질적 수요가 높은 온도, 압력, 배합비, 촉매 종류, 수율 지표 등 지능형 R&D의 핵심 변수를 중심으로 심층 큐레이션 수행
- 데이터 유효성 확보 : 추출된 데이터가 실제 연구개발의 입력값과 결과값으로 기능할 수 있도록, 도메인 지식을 기반으로 데이터 간의 인과관계를 매핑하여 자산 가치를 극대화

[세부전략1-2] 상호운용성 보장을 위한 표준화 및 품질 관리 체계 구축

- 글로벌 표준 스키마 수립 : 반도체, 바이오 등 주요 국가 전략기술 분야의 글로벌 표준과 호환되는 표준 데이터 스키마를 중심으로 수립하여, 구축된 데이터가 국내외 인프라와 즉시 연동될 수 있는 상호운용성 확보
- 단계별 품질보증(QA) 프로세스 : 알고리즘을 통한 자동 검증, 데이터 간의 논리 정합성 체크, 변리사 및 박사급 연구원의 최종 검증으로 이어지는 3단계 QA를 정립하여 데이터 무결성과 연구 재현성을 99% 이상으로 유지함
- 공유 및 활용 가이드라인 : 구축된 데이터 자산이 국가적 자산으로 지속 기능할 수 있도록 메타데이터 관리 규정을 수립하고, 수요기관 간의 안전한 데이터 연계를 위한 보안 및 비식별화 가이드라인 확정

□ [추진전략2] AI 기반 지능형 R&D 혁신 플랫폼 고도화

- (추진목표) 분절된 기술 지식을 통합하여 연구 난제를 해결하는 처방형(Prescriptive) R&D 모델 구현

[세부전략2-1] 다차원 관계 분석 기반의 지능형 추론 로직 도출(IP-AI R&D 모델)

- 고정밀 예측 알고리즘 개발 : 큐레이션된 고품질 IP 데이터를 활용하여 기술 간 상관관계를 분석하고, 적은 양의 전문 데이터로도 높은 예측 정밀도를 확보할 수 있는 지능형 추론 알고리즘을 개발함
- 단계적 유효성 검증 : 1단계(9개월)에서는 확보된 자산의 가치를 입증하기 위한 핵심 알고리즘의 유효성 검증(PoC) 및 Lab 수준의 시뮬레이션을 수행하여 지능형 R&D 방법론의 기술적 타당성을 조기에 확립함
- 모델 최적화 및 경량화 : 산업 현장의 다양한 컴퓨팅 환경을 고려하여, 분석의 신속성을 확보하기 위한 모델 경량화 및 도메인 간 데이터 호환성을 체계화하여 솔루션의 범용성을 높임

[세부전략2-2] 연구 전주기를 지원하는 지능형 R&D 통합 플랫폼 구현

- 처방형(Prescriptive) 솔루션 플랫폼 : 연구자가 목표 성능을 입력하면 AI가 지식 자산을 기반으로 최적 실험 레시피(Recipe)나 설계안을 제안하는 지능형 어시스턴트 기능을 플랫폼에 탑재하여 연구 생산성을 혁신함

- **유연한 데이터 연계 아키텍처** : 신규 유입되는 글로벌 기술 데이터를 지속적으로 수용하고 통합할 수 있는 시스템 구조를 설계하여, 연구자가 항상 최신의 기술 트렌드가 반영된 처방을 받을 수 있는 환경을 조성함
- **사용자 친화적 UI/UX** : 연구 현장의 실제 워크플로우(기획→실험→분석)를 반영한 인터페이스를 구축하여, 연구원이 별도의 코딩 지식 없이도 지능형 솔루션을 R&D에 즉시 활용할 수 있도록 지원함

□ [추진전략3] 기술 가치사슬 확장 및 실증 가속화

- **(추진목표)** 단위 기술 개발뿐만 아니라 산업 현장의 실질적 매출 성과 및 생산성 혁신 입증

[세부전략3-1] 데이터 스펙트럼 확장을 통해 기술 및 제품군 단위의 통합 자산화

- **가치사슬 통합 데이터 레이크 구축** : 1단계 단위 기술 중심의 검증 성과를 확장하여, 2단계에서는 전·후방 연계 기술군을 아우르는 통합 데이터 레이크를 구축하고 기술 간 융복합 시너지를 창출함
- **지속적인 자산(DB) 업데이트** : 새로운 특허 및 실험 데이터 도출 시 이를 지속적으로 추출하고 기존 지식 스키마에 자동 통합함으로써, 연구개발 전 주기에 걸친 지능형 자산 관리 체계를 완성함

[세부전략3-2] 기업 수요 연계 실증을 통해 산업 현장 안착 및 조기 상용화

- **실전형 테스트베드 운영** : 수요기업(제약, 제조 등)의 실제 제조 공정이나 소규모 테스트베드와 연계하여 시제품 개발 또는 현장 실증을 추진하고 솔루션의 현장 적합성을 최종 검증함
- **정량적 성과 입증** : IP-AI 솔루션 적용 전·후의 데이터를 정밀 비교하여 R&D 기간 및 비용 절감, 사업화 매출 성과 확대 등 구체적인 경제적 효용 입증
- **성과 기반 단계별 관리** : 단계 평가를 통해 성과가 우수한 과제를 엄선하여 후속 지원하는 성과 환류 체계를 운영하며, 실증 과정에서 도출된 현장 데이터를 모델에 다시 피드백하여 알고리즘의 정밀도를 지속적으로 고도화함

제4장 기대효과

제1절 기술적 효과

IP 중심 R&D의 데이터·AI 기반 혁신 전환과 IP 데이터 스케일업을 통한 산업문제 해결 및 상생형 R&D 생태계 구축으로 국가 지식재산 역량 강화

- (지능형 기술 정보 분석 기술의 내재화) 해외 솔루션 의존도가 높은 고난도 기술 문헌 분석 및 정보 추출 분야의 독자적 기술력 확보
 - 단순 단어 매칭 방식의 한계를 넘어, 복잡한 기술 도면과 서술형 문서의 맥락을 스스로 이해하고 구조화하는 핵심 분석 엔진을 국내 자체 개발
 - 파편화된 기술 정보 간의 상관관계를 자동으로 매핑하는 고신뢰 추론 기술 확보로 글로벌 지식 정보 시장의 기술 종속 탈피
 - 데이터 수집부터 지능형 분석까지 연결되는 기술 정보 선순환 체계의 국내 완성을 통해 국가 차원의 기술 주권 확립
- (데이터 중심의 연구 의사결정 지원 고도화) 대규모 언어모델과 전문 지식 데이터셋을 결합한 최적 경로 탐색 알고리즘을 개발하여, 기존 인력 중심 연구 기획의 한계를 기술적으로 극복
 - 연구 현장에서 발생하는 기술적 난제에 대해 실시간으로 최적의 해결 방안을 제시하고 의사결정을 지원하는 지능형 시스템의 핵심 엔진으로 활용
 - 기술-시장-특허를 연계한 다차원 통합 분석 기술과 결합하여 성공 가능성이 높은 유망 기술을 조기에 발굴하는 선제적 연구 환경 구축
- (기술 정보 관리의 글로벌 표준 선도) 국내 연구 현장에 최적화된 데이터 관리 규격 및 기계 학습용 가이드라인을 수립하여, 국내 산업계의 글로벌 기술 경쟁력 확보 기반 마련
 - 자체 기술 분석 표준 확보를 통해 주요국의 지식재산 관련 기술 규제에 선제적 대응이 가능하며, 해외 선진국 중심의 정보 독점에 대한 기술장벽 위험 해소

제2절 경제적 효과

고정밀 IP 데이터 자산화와 AI 융복합을 통한 국가 전략기술 경쟁력을 강화하고, 가이드 보급을 통한 연구개발 자원의 효율적 운용

- (연구개발 자원의 효율적 배분 및 손실 방지) 지능형 정보 분석 가이드 보급을 통한 직접적 연구 비용 및 수행 기간의 획기적 저감
 - (중복 투자 피해 경감) 유사 기술의 미확보로 인해 발생하는 불필요한 연구 예산 낭비를 사전 차단하여, 국가 전체 R&D 자원의 효율적 운용 및 경제적 손실 완화
 - 기술적 사각지대 파악 미비로 인한 특허 분쟁 비용 등 막대한 사회적 비용을 사전에 경감하고 기업의 경영 안정성 강화
 - 데이터 기반의 정밀한 연구 기획으로 기술 사업화 성공률을 높이고 선도적 시장 진입을 통한 매출 증대 효과 기대
- (국가 핵심 산업의 지속가능한 성장 동력 보호) 지능형 기술 분석 플랫폼의 현장 적용으로 국가 전략 산업의 유기적 가치 보호
 - (연구 전주기 생산성 제고) 고난도 비정형 문서 분석의 소요 인력과 시간을 획기적으로 줄여 산업계 연구 생산성 향상 효과 기대
 - 첨단 제조 및 바이오 등 미래 핵심 분야의 기술 트렌드를 실시간 탐지하여 유망 시장 선점 및 국가 성장 잠재력 보호
 - 정보 접근성이 취약한 중소·중견기업에 지능형 분석 솔루션을 보급함으로써 산업 전반의 기술 기획 역량 상향 평준화 견인
- (신개념 지식 서비스 산업의 활성화) 기술과 데이터가 융합된 새로운 비즈니스 생태계 확대 및 고부가가치 일자리 창출
 - 글로벌 데이터 서비스 시장 성장에 발맞춰 국내 기업의 데이터 처리 경쟁력을 확보하고 해외 지식 서비스 시장 진출 기반 마련
 - 원천 기술의 민간 이전을 통해 지식 서비스 스타트업 창업 및 기존 기업의 신사업 진출을 유도하여 데이터 중심의 산업 생태계 활성화 효과 기대

제3절 사회적 효과

**IP-AI 연계형 핵심 플랫폼을 통해 산업 밸류체인
동반 성장을 지원하고, 공정한 기술 거래 및 활용을
촉진하여 글로벌 기술 경쟁력 강화에 기여**

□ R&D 생산성 혁신 및 IP 데이터 기반의 산업 혁신 생태계 다각화

- AI 기반의 처방형(Prescriptive) 솔루션을 연구 현장에 적용하여 R&D 시행착오를 획기적으로 단축(기간 및 비용 절감)하고 연구개발 생산성 극대화
 - 소재 역설계, 표준특허 자동 창출 등 IP-AI 연계형 핵심 플랫폼을 통해 기존 R&D의 한계를 극복하는 신시장 패러다임 창출 및 시장 진입 가속화
- 개별 기업의 단위 기술 확보를 넘어, 전주기 공정 통합 실증 및 상용화를 지원함으로써 산업 밸류체인 구성원 전체의 동반 성장과 국가 지식재산 총역량의 비약적 향상 지원

□ (데이터 주도의 투명한 연구 질서 확립) 객관적 정보 기반의 연구 환경 조성을 통해 건전하고 공정한 혁신 생태계 강화

- (기술 정보 접근성 제고) 고도화된 지식 분석 기술의 대중화를 통해 연구자가 최신 정보를 명확히 파악할 수 있는 연구 환경 조성 및 확산에 기여
 - 누구나 쉽게 활용이 가능한 지능형 분석 도구를 보급하여 연구자 간의 정보 격차를 기술적으로 해소하고 상생하는 혁신 생태계 구축
 - 지식 자산의 가치를 명확히 평가하는 체계를 마련하여 공정한 기술 거래 및 활용을 촉진하고 정당한 기업 활동을 보호하는 환경 조성

□ (국가 대외 위상 및 기술 경쟁력 강화) 한국형 기술 정보 관리 체계의 우수성을 대외적으로 입증함으로써 글로벌 기술 경쟁력 강화 기여

- 지식 자산 관리 미비로 인한 국가적 기술 경쟁력 훼손을 방지하고, 연구 성과에 대한 글로벌 시장의 신뢰도를 기술적 수단으로 보장
- 중소 연구소 및 기업의 데이터 역량을 국가 차원에서 상향 지원하고, 국가 핵심 지식 자산의 체계적 전환으로 기술 주도권 확보

제 5장 결론

□ R&D 패러다임의 데이터·AI 중심 전환 가속화

- R&D 방식이 기존 경험·이론 기반 반복 실험에서 데이터·AI 기반 가설 설정 및 결과 예측 중심으로 빠르게 전환되면서, 고비용·저효율 구조를 탈피한 초고속·저비용 혁신 R&D가 새로운 표준으로 부상
- 글로벌 기업과 주요국은 AI를 활용해 연구기간을 획기적으로 단축하고 있으며, 데이터 자산화를 통해 신소재·신기술 개발 등에서 파괴적 혁신 성과를 창출하는 등 기술 패권 경쟁이 심화

□ 특허데이터 기반 R&D 혁신 자산 가치 및 전략성 증대

- 특허는 전 세계 기술정보의 약 80%를 포함하는 핵심 데이터 자산으로, 기술 공백·경쟁력·분쟁 위험 등을 예측할 수 있는 유일한 전략적 정보원으로 기능
- 6억 건 이상의 글로벌 특허 데이터는 R&D 방향 설정과 기술 전략 수립의 핵심 기반이며, 권리 소멸 분석 등을 통해 리스크 최소화 및 고효율 기술개발이 가능
- 이에 따라 IP 전주기(창출 - 활용 - 보호 - 기반)에 걸친 데이터 기반 혁신 지원체계 구축 필요성이 증대

□ AI-Ready 데이터 기반 R&D 인프라 구축 시급

- AI 중심 R&D 환경에서는 성과의 핵심이 인력·자본이 아닌 AI 활용 역량과 고품질 데이터 확보로 이동하고 있으며, AI가 즉시 학습 가능한 AI-Ready 데이터 구축이 핵심 과제로 부상
- 국가 R&D 성과물 다수가 비정형·비라벨링 상태로 존재하여 AI 활용에 한계가 있으며, 데이터 구조화 및 기계가독성 확보 없이는 생산성 혁신 실현이 어려운 상황
- 이에 따라 국가 R&D 성과를 AI 학습 가능한 형태로 전환하는 데이터 인프라 고도화가 시급

□ IP 데이터 기반 R&D 효율화 체계 혁신 필요

- IP-R&D는 기술료 4.4배 등 높은 성과에도 불구하고 수작업 중심 분석 구조로 인해 확산에 한계가 존재하며, 분석 프로세스의 자동화·데이터 기반 전환 필요
- 특허 빅데이터와 AI 융합을 통해 기술 난제 해결, 중복투자 방지, 핵심특허 확보 등 R&D 전 과정의 효율성을 제고할 수 있는 체계적 혁신 필요
- 이를 통해 지식재산 데이터가 단순 분석 도구를 넘어 국가 R&D 생산성 혁신을 견인하는 핵심 인프라로 기능할 필요성 증대