

http://www.bioin.or.kr

BioInGlobal

바이오인글로벌

BioInGlobal

글로벌동향 보고서 [No.67]

(전문가) 김경남

글로벌 통신원 (KIST Europe 방문연구원)

CONTENTS

새로운 바이오의료 R&D
패러다임 : 자가학습 및
AI 활용



BioInGlobal
No.67
(2022-12)

[BioInGlobal No.67]

새로운 바이오의료 R&D 패러다임 : 자가학습 및 AI 활용

작성자 : 김경남 (KIST Europe 방문연구원)

출처① Mckinsey. 2022. 7., ② Mckinsey. 2022. 10.

※ 본 원고는 원본 원고의 내용을 증개하여 각색한 내용을 포함하고 있습니다.

(원제① : Self-learning: The dawn of a new biomedical R&D paradigm)

<https://www.mckinsey.com/industries/life-sciences/our-insights/self-learning-the-dawn-of-a-new-biomedical-r-and-d-paradigm>

(원제② : AI in biopharma research: A time to focus and scale)

<https://www.mckinsey.com/industries/life-sciences/our-insights/ai-in-biopharma-research-a-time-to-focus-and-scale>

【요약】

- 본 보고서는 바이오 분야 디지털 전환 및 AI 활용에 대한 Mckinsey&Company의 최신 보고서 2편을 정리하였음
- 보고서 I. 자가학습 : 새로운 바이오메디컬R&D 패러다임
 - 과학과 기술은 바이오의료 R&D 가치 사슬의 각 단계에서 혁신을 주도하며, 과학-기술 간의 피드백 루프 강화를 통해 환자에게 도움이 되는 자가 학습 모델을 촉발할 수 있음
 - 본 보고서에서는 바이오의료 R&D에서 새롭게 부상하는 솔루션과 접근방식으로 ①질병 이해, ②치료가설 생성, ③치료 모달리티 혁신, ④in silico 및 in vitro 검증 방법, ⑤임상 및 실제 증거 피드백의 5가지 요소를 식별하여 새로운 바이오의료 R&D 패러다임 개념 및 사용 사례를 제시함
- 보고서 II. 바이오 제약 연구에서의 AI 활용
 - 바이오 제약기업들은 특정한 과학적 문제 및 운영 상의 문제점에 집중하여, AI를 연구 업무흐름에 완전히 통합함으로써 환자들에 더 큰 영향력과 함께 중요 가치를 제공할 수 있음
 - 바이오 제약 기업이 AI 기반 신약발굴 기업과 파트너십의 효과를 극대화하려면 두 가지 잠재적 장애물*을 인식하여 바이오 제약기업들은 내부 역량 구축과 AI

기본 신약 개발 회사와의 파트너십 간에 균형을 유지해야 함

- * (1) AI 기반 신약개발 방식(파트너십 포함)은 기존 R&D 활동과 거리가 있으며 바이오 제약기업의 과학 및 운영 프로세스에 고정되지 않음
- (2) 내부 R&D 팀 내 디지털화된 약물 발견 기능 및 데이터에 대한 투자는 관련 기술 및 기능 구축보다 파트너 플랫폼 활용 및 IP 강화에 더욱 많이 활용되고 있음

- 바이오 제약 연구에서 AI를 성공적으로 구현하기 위해 다음 네 가지 영역이 중요

- ① 전략 및 설계 기반 로드맵 작성
- ② 분기별 가치 발표(QVR; quarterly value releases)에 중점을 둔 끊임없는 가치 제공
- ③ 파트너십을 통한 IP 및 역량 구축, 번역 전문성 개발
- ④ MLOps(머신러닝 운영) 및 재사용 가능한 분석 자산을 통한 AI의 산업화

[목차]

[보고서 I] 자가학습 : 새로운 바이오메디컬R&D 패러다임

1. 개요
2. 바이오의료 R&D에서 새롭게 부상하는 솔루션과 접근방식
3. 새로운 바이오의료 R&D 패러다임
4. 새로운 바이오의료 연구개발 패러다임을 위한 새로운 조직 모델
5. 결론

[보고서 II] 바이오 제약 연구에서의 AI 활용

1. 개요
2. AI 기반 신약 개발 산업
3. 기업 내부에서 전사적 역량을 구축하는 데 따르는 실질적인 영향
4. 바이오 제약 연구에서 AI를 성공적으로 구현하려면 무엇이 필요합니까?
5. 결론

[보고서 I] 자가학습 : 새로운 바이오메디컬R&D 패러다임

1. 개요

- 세포 및 유전자 치료법의 발전(예: 키메라 항원 수용체 T 세포 및 mRNA 기반 치료법) 질병 치료를 위해 면밀히 설계되고 표적화된 개입의 잠재력을 보여주는 사례임, 이 잠재력은 최근까지 상상할 수 없었던 속도로 수많은 질병을 치료하고 예방하기 위한 정확한 치료법 설계에 활용될 수 있음
- 이러한 발전을 가능하게 하는 원동력은 데이터, 자동화, 컴퓨팅 성능 및 AI 활용의 발전과 함께 생물학의 혁신이 결합된 것임. 이는 이미 발굴 단계에서 임상시험, 실제 사용에 이르기까지 바이오의료 R&D 가치 사슬의 각 단계에서 개발을 주도하고 있음
- 그러나 프로세스의 각 단계에서 생성된 데이터와 통찰력을 연결할 수 있는 자동화 및 AI를 도입하면 완전히 새로운 R&D 패러다임이 시작될 수 있음

- 바이오의료 R&D에서 오늘날의 환원주의적 접근*은 과거의 방식이 될 것임
- * 과학자들이 새로운 치료법을 개발하고 검증하기 전에 질병 생물학의 단일 구성 요소 또는 기능에 확대하고 시행 착오를 통해 (종종 잘못 정의된 표현형 질병 상태에 대해) 세심하게 가설을 테스트하는 방식
- 대신, 훨씬 더 정확한 질병 상태에 대해 R&D 가치 사슬의 다양한 단계 사이에서 데이터 피드백 루프에 의해 추진되는 순환 프로세스의 성공을 위해 요법이 체계적으로 설계될 것임. 하나의 단계에서 얻은 데이터와 통찰력은 가치 사슬의 상/하단에 있는 다른 사람들에게 정보를 제공하고 다른 질병에 대한 이해도 강화할 것임

- 이를 통해 기하급수적 약물 혁신이 이루어질 수 있음. 실패한 임상 후보는 줄어들고 새롭고 매우 효과적이며 안전한 치료법은 더 많아질 것임. 말기 증상뿐만 아니라 질병의 초기 신호에 반응하는 예방 및 개입이 있을 것이며, 질병 치료는 다른 하위 집단을 주의 깊게 표적화할 것임
- 우리(Mckinsey)는 세계가 “바이오 레볼루션 Bio Revolution*” 패러다임을 가능케 하는 변곡점에 가까워지고 있음을 제시함. 본 보고서는 R&D 가치 사슬의 각 단계에서 바이오 혁명을 향해 나아가고 있는 전 세계 실험실에서 진행 중인 기술 기반 연구 혁신 사례 중 일부를 조사하였음. 또한 가치 사슬 전반에 걸쳐 데이터와 통찰력을 통합하여 발전을 가속화하기 위해 바이오제약 기업들이 무엇을 할 수 있는지를 조사하였음

* 바이오 혁명에 대한 맥킨지 보고서(2020) 참조:

<https://www.mckinsey.com/industries/life-sciences/our-insights/the-bio-revolution-innovations-transforming-economies-societies-and-our-lives>

- 바이오 제약 기업들이 이러한 연결을 만들기 위해서는 새로운 데이터와 기술 인프라가 요구됨. 또한 새로운 R&D 패러다임은 지금보다 훨씬 더 많은 협력을 요구하기 때문에 연구자와 기계뿐만 아니라 연구자들 자신과 외부 파트너간 협업을 요구하기 때문에 현재와 다른 조직운영을 고려해야 함

2. 바이오의료 R&D에서 새롭게 부상하는 솔루션과 접근방식

- 새로운 바이오의료 R&D 가치 사슬의 가장 주목받는 발전을 고려하여 ①질병 이해, ②치료가설 생성, ③치료 모달리티 혁신, ④in silico 및 in vitro 검증 방법, ⑤임상 및 실제 증거 피드백의 5가지 요소를 식별하였음
- 상기 요소들은 현재 진행 중인 발전의 정도를 나타낼 수 있음. 일부는 잘 확립되어 있는 반면, 아직 널리 채택되지 않았거나 완전히 검증되지 않았으며, 어떤 것이 가장 큰 영향을 미칠지 아직 명확하지 않은 요소들도 존재함. 다만, 이 요소들의 집단적 힘은 논쟁의 여지가 없음 (BOX-1 “새로운 바이오의료 R&D의 5가지 요소: 신호 및 조력자” 참조).

[BOX-1] 새로운 바이오의료 R&D의 5가지 요소 : 신호 및 조력자(enablers)

- 신기술과 생물학적 혁신의 융합은 생물학 R&D 가치 사슬의 모든 요소에 걸쳐 점진적인 변화를 주도함. 이러한 새로운 접근 방식 중 일부는 이미 업계 선도기업들이 채택하고 있고, 다른 일부는 학계 및 생명공학 신생 기업 전반에 걸쳐 잠재적인 사용 사례로 구체화되고 있으며, 일부는 실현되지 않거나 아직 가시화되지 않은 상황임
- 다음은 완전히 구체화되었을 때 바이오의료 산업을 R&D 가치 사슬의 단계적 변화와 재구상에 더 가깝게 만들 수 있는 일련의 선택 신호 및 인에이블러(새로운 혁신 아이디어가 풍부함에 따라 완전하지 않음)임. 이러한 혁신과 다른 혁신이 결합되면 보다 상호 연결되고 기술이 지원되는 5개 요소의 생물 의학 R&D 패러다임을 생성할 수 있음

① 질병 이해 Disease understanding

- **Omics 기반 질병 통찰력 Omics-driven disease insights** : Multiomic 데이터 세트는 유전자-질병 관계를 보다 강력하게 설명해줄 수 있는 초대형 게놈 데이터 뱅크를 통해 진단 및 치료 접근 방식을 효과적으로 알리기 시작. 또한 장기 및 질병의 구조적/기능적 매핑은 새롭고 신뢰할 수 있는 바이오마커 및 질병 모델로 이어지게 됨
- **높은 처리량의 데이터 수집 및 통찰력 생성 High-throughput data collection and insight generation** : 환자 유래 유도 만능 줄기 세포를 기반으로 하는 질병 모델, 고처리량 세포 페인팅 분석, 차세대 현미경 기술(예: 극저온 전자 단층 촬영)과 같은 새로운 실험 접근 방식은 건강 및 질병 상태에 대해 신속한 대규모 병렬 비교를 가능하게 함
- **개선된 질병 분류법 Improved disease taxonomies** : 건강과 질병은 연속체로 간주되며, 증상에서 근본적인 원인으로 포커스가 이동됨
- **인실리코 합성생물학 모델링 In silico system biology modeling** : 기술은 개별 경로, 세포 및 기관의 시뮬레이션을 허용하고 최종적으로 전체 유기체의 시뮬레이션을 모델링함

② 치료 가설 생성 Therapeutic-hypothesis generation

- **향상된 기본 데이터 Better underlying data** : 실제 또는 예측된 구조 생물학 데이터의 가용성이 높아짐에 따라 새로운 분자 표적을 쉽게 탐색할 수 있음. 또한 사용 가능한 데이터의 깊이가 증가하고 지식 그래프가 향상되어 보다 정교한 가설 설정이 가능
- **기술 조력자 Tech enablers** : 보다 전통적인 계산 능력은 복잡한 in silico 모델링의 가능성을 높이고 새로운 알고리즘은 데이터를 조사하고 통찰력을 도출하는 새로운 방법을 제공함. 양자 컴퓨팅과 같은 새로운 접근 방식은 한 번에 더 넓은 범위의 구조 옵션 또는 분자-분자 상호 작용을 탐색하는 데 용이함
- **더 많은 방법과 응용 프로그램 More methods and applications** : 딥 러닝 지원 분자 도킹(deep-learning-assisted molecular docking)과 같은 방법들의 레퍼토리가 증가함에 따라 생성된 가설의 다양성이 증가함. 대규모 용도 변경 및 생물학 기반의 높은 처리량 대상 식별과 같은 가설 자동화는 잠재적으로 초기 유입 경로를 더욱 확장하게 됨

③ 치료 모달리티 혁신 Therapeutic-modality innovation

- **재생 요법 Regenerative therapies** : 연구자들은 각 환자, 적응증, 장기 또는 조직에 적합한 줄기 세포를 식별하고 그에 따라 최적의 결과를 위해 분화를 활성화하는 방법을 학습함. Bioactive scaffold는 세포외 복구 메커니즘을 활성화함
- **세포 및 조직 공학 Cell and tissue engineering** : 유전자 변형 세포 치료법과 엔지니어링된 면역 세포는 보다 일상적으로 치료 및 진단방법으로 채택됨. 조직, 장기 및 연골을 프린팅하는 능력은 이식에 대한 새로운 지평을 열어줌
- **유전자 코드 디자인 플랫폼 Genetic code design platforms** : 메가 뉴클레아제 기반, TALEN, CRISPR-Cas9, CRISPR-Cpf1 및 핵산 기반(예: 안티센스 및 mRNA 기반) 모달리티와 같은 새로운 유전자 치료 및 유전자 편집 플랫폼, 치료 및 연구 도구 키트 확장
- **개선된 전달, 내구성 및 면역원성 Improved delivery, durability, and immunogenicity** : 차세대 유전자 치료법은 향상된 전달, 내구성 및 조절 기능을 제공함. 면역원성을 예측하고 적응시키는 능력은 항체, 세포 및 유전자 요법의 다양성을 증가시킴. 예를 들어, 위장관의 선택적 코팅은 표적화된 경구 전달이나 약물의 연장된 반감기를 허용
- **표적 및 다기능 모달리티 Targeted and multifunctional modalities** : 단백질 분해 표적 키메라, 자가포식 표적 키메라, 리소좀 표적 키메라 및 분자 접착제 분해제와 같은 알려진 분자 표적에 대한 새로운 접근 방식은 치료 레퍼토리를 확장함. 다중특이성, 최소화 및 접합 항체를 포함한 차세대 항체는 생물학적 치료제의 표적화 및 안전성을 개선함
- **향상된 조합 Better combinations** : 시너지를 가진 약물 조합은 문헌 및 멀티오믹스 데이터를 기반으로 인실리코(in silico)로 예측됨. 수술, 방사선, 정신/물리 치료, 전기 자극, 가상 현실, 새로운 치료법의 새로운 조합은 환자의 상태를 개선하는 데 도움이 될 것임
- **미생물 공학 및 합성 생물학 Microorganism engineering and synthetic biology** : 생물학적 제조는 보다 지속가능하고 다양한 생산을 가능케 함. 생체 분자 응축물과 같이 잘 알려지지 않은 세포 현상에 대한 탐구는 치료적 개입의 새로운 길을 제시함. 자가 증폭 또는 리프로그래밍이 가능한 분자는 mRNA백신 접종으로 지속적인 면역을 가능하게 함. 미생물군집은 합성 미생물을 통해 정밀하게 조절되어 질병에

대응하거나 예방할 수 있음

④ In silico 및 in vitro 검증 방법 In silico and in vitro validation methods

- **In silico 우선 최적화 In-silico-first optimization** : 분자 특성과 생물학적 활성 모두에 대한 in silico 모델링을 예측하면 개발 위험성을 낮추고 가장 관련성이 높은 실험들의 우선순위 결정에 도움
- **개선된 인간 생물학 기반 검증 모델 Improved human-biology-based validation models** : 세포 온 칩, 장기 온 칩 또는 환자 온 칩을 사용하는 신속한 테스트 모델은 개별 환자의 유전 또는 단백질 구성 또는 질병의 세포 환경을 복제함. 오가노이드는 단일 세포에서 일어나는 일을 시뮬레이션할 뿐만 아니라 인간 장기의 전체 3D 환경을 재현할 수 있음
- **바이오마커, 바이오센서 및 분석 Biomarkers, biosensors, and assays** : 바이오센서, 분석 및 바이오마커 플랫폼의 레퍼토리가 증가함에 따라 시험관 내 검증 속도가 빨라짐
- **엔드 투 엔드 자동화 End-to-end automation** : 완벽하게 통합된 AI 기반 로봇공학은 완전 자동화된 초고처리량 스크리닝 및 연속적인 실험을 동시에 가능케 함

⑤ 임상 및 실제 증거 피드백 Clinical and real-world evidence feedback

- **모의 실험 Simulated trials** : 사전 시뮬레이션은 임상 시험에서 부작용의 위험을 예측하고 각 환자에 대한 최상의 반응자와 이상적인 치료 계획을 식별함
- **데이터 수집 및 사용 개선 Better collection and use of data** : 기존 디지털 바이오마커(실시간 바이오마커 포함)의 개선된 사용은 질병 예방을 가능하게 하고 질병 치료 결과에 대한 데이터를 강화하여 연구 노력에 더 나은 정보를 제공함. 또한 자연어 처리를 통한 비정형 의료 데이터 마이닝은 치료, 증상, 결과 및 충족되지 않은 요구 간의 연결에 대한 이해를 심화시켜줌. AI는 이미 시장에 나와 있는 약품을 재검토하여 치료 표준을 강화하고 새로운 잠재적 응용 프로그램을 탐색하는 데 도움이 됨
- **정확한 치료 설계 Precise treatment design** : 신속한 시퀀싱과 같은

새로운 기술을 활용하는 정밀 진단은 질병 분류 및 치료 패러다임을 개선하는 데 도움을 줌. 향후 미래에는 맞춤형 개별 환자 치료 계획이 개별화된 오가노이드 또는 칩에 대해 사전 평가될 것으로 전망됨

① 질병 이해 Disease understanding

- 질병 치료 및 예방을 위한 더 많은, 더 나은 치료법 개발의 빠른 발전을 방해하는 주요 요인은 질병의 모든 다양한 징후와 건강의 기초가 되는 메커니즘에 대한 이해가 제한적이라는 것임. 20년 전 인간 게놈 지도 작성은 인간 생물학에서 많은 새로운 연구 길을 열어주는 중요한 진전이었으나, 유전자는 건강과 질병이라는 더 넓은 퍼즐의 일부일 뿐이며 대부분의 질병을 해결할 수 있을 만큼 충분히 완전한 그림을 제공하지 못하였음
- 오늘날 중요한 발전에는 AI가 분석할 수 있는 방대한 체외 데이터 생성을 위한 세포 페인팅(cell painting), 전체 인구의 다중오믹스 측정(특히 전사체 및 단백질체학), 익명화된 전자 건강 기록과 같은 새로운 실험적 접근 방식이 포함됨. 이는 의료 산업이 동반 질환, 질병진행 및 개인 간의 차이를 고려하여 인간의 건강 상태와 질병 상태를 시뮬레이션하고 더 정확하게 정의하는 데 도움이 될 수 있음

② 치료 가설 생성 Therapeutic-hypothesis generation

- 질병에 대한 보다 총체적이고 데이터 중심적인 이해는 체계적이고 확장 가능한 치료 가설 생성을 가능하게 함. 오늘날 과학자들은 돌파구를 찾기 위해 특정질병 또는 바이오마커와 관련된 개별 세포 유형 또는 경로를 탐색하는 경향이 있음. 세 가지 측면의 진전된 진행 상황은 이 같은 경향을 변경하여 질병과 관련된 이전에 알려지지 않은 생물학적 상호의존성을 발굴하고 가설을 빠르게 생성하기 위한 데이터의 신속한 탐색을 촉진할 것임
- **더 많은 데이터에 대한 더 나은 액세스 Better access to more data.** 더 많은 양의 질병 데이터와 더 다양한 질병 데이터 뿐만 아니라 이러한 데이터에 대한 접근성도 개선되었음. 이러한 이유는 오픈 액세스 데이터베이스의 등장에 따른 것으로, 예를 들어 게놈 데이터, 생체 분자의 구조 및 기능 데이터, 스크리닝 데이터는 모두 오픈 액세스 데이터베이스에서 사용할 수 있음. 이러한 데이터를 주의 깊게 사용하면 과학자들이 기존 약물을 알려진 표적에 대한 용도 변경 및 새로운 약물 설계에 대한 가설을 테스트하는 데 도움이 될 수 있음

· **기술 인에이블러 Tech enablers.** 값싸고 풍부한 컴퓨팅 성능, 양자 컴퓨팅의 출현 및 기계 학습 방법은 생물학 분야에서 점점 더 복잡해지는 분석 작업을 해결하는 데 도움이 되는 도구 중 하나임

· **in silico 가설 생성 자동화 Automation of in silico hypothesis generation.** in silico 가설의 생성을 자동화하면 질병과 경로 사이뿐만 아니라 질병과 유전자, 영양 및 행동과 같은 다른 요인 호스트 사이에서 이전에 고려되지 않은 상관 관계에 대한 고처리량 탐색이 용이해지게 됨. 또한 가설의 편향을 없애고 기 확립된 과학 분야에서 사용되는 가설을 개선하는 데 도움이 될 수 있음

③ 치료 모달리티 혁신 Therapeutic-modality innovation

- 질병에 대한 더 나은 이해와 과학 기술의 발전은 질병이나 환자에게 특별히 맞춤화된 코드와 같은 치료법으로 이어질 수 있음. 예를 들어, CRISPR-Cas9, mRNA 및RNAi를 포함한 여러 새로운 모달리티 플랫폼은 유전자 코드를 표적으로 하고 있음. 이러한 유형의 치료제는 생물학적 문제를 생물학적 모델 또는 약물 후보로 신속하게 변환하여 시작을 가속화할 수 있다는 추가 이점이 있음
- mRNA의 선형적이고, 코드와 같은 서열 때문에, 예를 들어 표적 항체 또는 소분자 억제제를 확인하고 합성하는 것보다 암에 미치는 영향을 테스트하기 위해 설계하고 합성하는 것이 더 용이함. 조작된 세포(예: 면역 및 재생 요법), 다기능 모달리티(예: 항체 약물 접합체 및 단백질 분해 표적 키메라) 및 합성 미생물(예: 장내 미생물 군집의 재조정)는 다른 많은 새로운 모달리티 중 하나임. 모달리티 혁신은 생물학적 제제에만 국한되지 않음. 예를 들어 향상된 계산 방법은 보다 정밀하게 설계된 소분자로 이어질 수 있음
- 한편, 재료과학 및 합성생물학과 같은 분야의 발전은 기존 모달리티(예: 더 나은 전달, 더 많은 내구성 또는 더 적은 면역원성을 통해)를 더욱 개선하거나 새로운 것을 개발하는 데 도움이 될 것임. 또한 가까운 미래에는 오늘날 의사들이 환자가 표준치료에 잘 반응하지 않을 때만 탐색하는 개인화된 중재 조합을 설계, 개발 및 테스트하는 것이 가능할 수도 있음. 이러한 조합(예 : AI 보조 수술 이후 약물 처방, 디지털 치료제, 미생물군집 이식 및 상태를 모니터링을 위한 웨어러블 장치 및 어플리케이션)은

시너지를 극대화하도록 면밀하게 설계될 것임

④ In silico 및 in vitro 검증 방법 In silico and in vitro validation methods

- 오늘날 과학자들은 환자의 유전적 구성을 복제하거나 질병의 세포 환경을 나타내는 신속하고 높은 처리량의 세포-온-칩 또는 장기-온-칩 테스트 모델을 생성할 수 있음. 이와 유사하게 오가노이드는 인간 장기의 3차원 환경을 재창조하여 종 간 모든 생물학적 차이를 고려할 수 없는 동물 모델을 사용한 결과나 장기의 넓은 환경을 고려하지 않는 세포주 표준화보다 더 잠재적으로 더 정확한 결과를 가져올 것임
- 과학자들이 체외모형을 통해 질병에 대해 더 많이 배울수록 이를 반영하는 예측가능한 in silico 모델 설계가 더 용이해짐. 과학자들이 분자 특성(예: 독성, 흡수, 분포, 대사 및 배설) 뿐만 아니라 면역원성과 약물-마이크로바이옴 상호작용도 예측하기 위해 in silico 모델을 활용할 수 있는 충분한 데이터를 갖게 될 때가 도래하게 될 것임. 시간이 지남에 따라 약물 후보에 대한 전임상 필터링은 동물 또는 체외 모델이 아닌 in silico 모델에서 더 많이 수행될 수 있어 처리량을 높이고 치료 개발과 관련된 위험을 낮출 수 있음

⑤ 임상 및 실제 증거 피드백 Clinical and real-world evidence feedback

- 기술은 대량의 데이터 생성 및 수집을 용이하게 함. 신약 후보물질의 임상을 통해 질병에 대한 데이터가 많이 축적될수록 미래의 가설 생성과 검증이 더욱 집중되고 정확해질 가능성이 높음. 전자 건강 기록에 수집된 데이터 및 기타 실제 데이터(예 : 24/7 데이터를 생성할 수 있는 바이오마커 및 웨어러블의 광범위한 측정)도 마찬가지이며, 이러한 측정은 보다 강력한 환자 특성화로 이어질 수 있음. 이는 보다 미묘한 질병 모델로 이어질 수 있음. 자연어 처리와 같은 계산 방법의 성숙은 새로운 데이터뿐만 아니라 문헌에서 구조화되지 않은 환자 데이터를 마이닝할 수 있음을 보장함

3. 새로운 바이오의료 R&D 패러다임

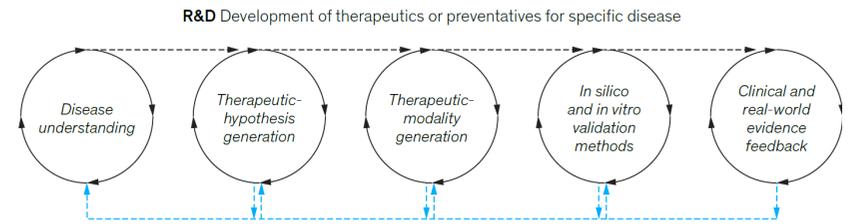
- 오늘날의 바이오의료 R&D 가치 사슬은 종종 선형적 가치 사슬로 표현되며, 이는 특정 질병에 대한 특정한 새로운 치료법을 추구하는 가치사슬의 한 단계에서 얻은 정보가 후속 단계에 어떻게 정보를 제공하는지를 앞방향을 향해 나타내는 일련의 갈매기형 표시와 함께 표현되고 있음

- 정보는 역방향으로도 흐르고 연구는 더 넓은 응용 분야를 가질 수 있음. 예를 들어 연구 가설을 검증하고 질병 생물학을 탐구하기 위한 다용도 도구인 CRISPR와 같은 플랫폼 기술의 출현이 그 예시임. 하지만 새로운 패러다임에서 프로세스는 더 넓은 목표를 가지고 있으며 과급되어짐
- 기술은 인간의 두뇌만으로는 감지하기 어려울 수 있는 새로운 패러다임 단계마다 통찰력을 발견할 뿐만 아니라 이들 간의 상호 의존성을 식별함. 또한 데이터와 통찰력이 오늘날보다 훨씬 더 자유롭고 빠르게 가치 사슬 상-하단으로 자동적으로 흐르도록 함. 이는 사람이 동의하고 각 반복을 시작하는 것보다 훨씬 더 반복적이고 순환적인 프로세스라 할 수 있음. 기존의 선형 R&D 프로세스는 더욱 상호연결된 프로세스(신속히 앞뒤로 정보를 공급하는 회전하는 바퀴형태의 일련의 프로세스)로 대체될 것임

그림 1: 새로운 바이오의료 R&D 패러다임

Future biomedical R&D will be an iterative process in which the insights from each step improve other cycles.

New biomedical R&D paradigm, illustrative



Learning Feedback-based iterations for every step improve other cycles and inspire new targets and application idea:

- 궁극적으로 새로운 패러다임의 목표는 캡처된 모든 데이터 포인트와 얻은 모든 통찰력을 단일 데이터 저장소에 공급하고 연결하는 것임. 알고리즘은 다양한 질병에 대한 이해와 치료를 개선하기 위해 데이터 저장소에서 끌어올 수 있음
- 특정 질병 상태를 재정의하지는 않더라도 더 나은 질병 이해는 훨씬 더 정확하고 확장 가능한 치료 가설로 이어지며, 이는 더 많은 고도로 맞춤형 치료 모달리티로 이어짐. 그 다음 이러한 가설에 대한 초기 테스트의 상당 부분을 in silico 및 in vitro에서 자동화할 수 있음

- 동시에 프로세스의 각 단계에서 얻은 교훈이 모든 이전 단계를 직접적으로 개선할 수 있기 때문에 정보의 역방향 흐름은 진행을 강화함. 체외 및 실제 세계에서 생성되고 in silico로 분석된 대용량 데이터는 질병 이해를 신속하게 알리고 새로운 가설을 생성하며 새로운 모달리티 플랫폼을 개발하는 데 도움이 됨
- 데이터에 대한 AI 평가는 정제된 실험 매개변수 또는 최적화된 치료 후보를 사용하여 또 다른 in vitro 테스트를 자동으로 제안할 수도 있음. AI는 또한 이러한 테스트의 실행을 시작할 수도 있음. 예를 들어, in vitro 테스트에서 약물 후보가 표적에 약한 결합력을 가지고 있음이 밝혀지면 AI는 약물 후보의 구조를 표적 구조와 비교하고 후보를 개선할 수 있는 몇 가지 방법을 생각해낸 다음 계속 진행할 수 있음. 가장 유망한 개선 사항을 선택하고, 실제 데이터를 기반으로 하는 시뮬레이션된 임상 시험에서 이를 종합하고 테스트함. AI 및 자동화의 광범위한 사용으로 새로운 R&D 가치 사슬은 의료 혁신을 가속화할 수 있음
(Box-2 “데이터 연결: 새로운 바이오의료 R&D 패러다임의 사용 사례“ 참조)

치매에 효과적인 약물 표적과 같은 가장 가능성이 높은 경로 단백질을 식별할 수 있다는 가설을 검증하거나 반증하는 가장 빠른 접근 방식으로 뇌 오가노이드 모델에서 일련의 표적 CRISPR 변형을 제안함

- 검증에 성공하면 분자 도킹 알고리즘은 다양한 형태의 치매를 치료하기 위해 합성되고 조사되는 우선순위 단백질에 대한 선택적 길항제에 대한 제안을 생성함. 이는 신경 퇴행성 질환 극복의 돌파구로 연결됨
- 이 접근 방식의 요소들은 이미 사용 중이며 이들 간의 피드백 루프를 강화하면 충족되지 않은 환자의 요구를 더 빠르고 더 적은 위험으로 해결하는 데 도움이 될 수 있는 자가 학습 모델을 만들 수 있음

4. 새로운 바이오의료 연구개발 패러다임을 위한 새로운 조직 모델

- 기업들은 이미 생물학적 진보, AI 및 자동화의 지원으로 새로운 바이오의료 R&D 가치사슬의 여러 요소를 연결하는 프로젝트를 진행하고 있음. 그러나 그 어떤 회사도 모든 데이터를 연결하여 어디서나 관련 데이터를 찾고 사용할 수 있는 시스템을 갖추고 있지는 않음. 이를 위해서는 데이터가 상호 연결되고 기계가 읽을 수 있으며 품질 개선 메커니즘(예: 거짓양성 식별)이 존재하는지 확인하기 위한 새로운 기술 인프라가 필요함. 또한 조직의 변화도 필요함
- 많은 조직에서 초기 R&D와 후기 R&D가 분리되어 유지되며 클리닉과 실제 세계의 증거는 연구자에게 천천히 피드백됨. 연구자들은 종종 단일 치료 영역에 초점을 맞춘 사업부에 고립되어 있으며 다양한 생물학 모델에 대한 많은 병렬 시스템과 분류법이 있는 경우가 다수임. 새로운 패러다임에서는 보다 연결된 R&D 프로세스를 반영하기 위해 이러한 경직된 조직체계를 완화해야 할 수도 있음. 이는 팀 구성 방식, 포함하는 기능 및 회사의 혁신 모델에 영향을 미칠 것임
- 팀은 빠른 정보 교환의 이점을 누리기 위해 훨씬 더 넓은 범위가 필요할 것임. 팀들이 R&D 프로세스의 모든 요소를 포괄하도록 역량이 확장된다면 팀들 간의 경계가 없어지게 될 것임. 그리고 팀에는 주제별로 깊은 전문

[BOX-2] 데이터 연결: 새로운 바이오의료 R&D 패러다임 사용 사례 (특정 심혈관 질환을 치료할 수 있는 신약 개발 상황을 가정)

- 다양한 세포 유형에서 다중오믹스를 측정하기 위한 In vitro 테스트(모든 약물 후보가 잠재적인 부작용을 조기에 예측하기 위한 일관적인 과정)는 하나의 약물 후보가 기존에 고려되지 않은 경로에서 여러 단백질의 발현 또는 활성을 하향 조절한다는 것을 보여줌
- 동시에 자연어 처리에 의한 웹 스크래핑은 이 동일한 경로가 기억을 포함한 인지 기능과 상관 관계가 있음을 보여줌. 알고리즘은 이러한 정보 조각을 찾아 연결하고 자동으로 스캔된 기존 문헌을 분석하여 현재 치매에 대한 이해가 이 경로와 치매로 이어지는 신경퇴행성 질환 사이의 직접적인 연관성을 식별하는지, 약물이나 치매 관련 증상을 완화하는 유전자 녹다운을 통해 이 단백질이 억제되는지 여부를 확인함
- 결과적으로 이러한 링크가 발견되지 않았을 경우, 임상시험 데이터는 구조적으로 유사한 단백질을 표적으로 하는 심혈관 질환에 대한 일반적인 표적 외 효과(예 : 두통과 같은)를 보고함. 이것은 추가적인 탐구를 촉구하고 AI는 새로운 경로가 조절되어 신경 퇴행에 대응하고

지식을 갖춘 전문가가 포함되지만 전체 가치 사슬을 이해하고 생물학과 기술의 잠재력을 활용할 수 있는 다분야 전문가도 필요함(예 : 높은 수준의 품질 데이터를 보장하는 가장 신뢰할 수 있는 과학적 접근 방식을 선택하는 전문가)

- R&D 팀이 신속하게 움직일 수 있도록 하려면 새로운 거버넌스 메커니즘이 필요할 것임. 예를 들어 팀은 유망한 치료 후보를 다음 단계로 발전시키고(자동으로 수행되지 않는 경우), 획기적인 혁신 잠재력을 가진 아이디어를 식별하고 우선순위를 지정하고, 예산 할당을 결정할 수 있는 권한이 필요함. 그리고 중앙 집중화된 의사 결정 프로세스는 자동화 및 AI로 인한 이점이 줄어들 수 있음. 또한 팀은 외부 전문 지식과 역량을 끌어낼 수 있는 권한이 필요함, 성공은 독점 약물, 기술 및 모달리티 플랫폼에만 의존하는 것이 아니라 알고리즘, 데이터 세트 및 디지털 솔루션에도 달려 있기 때문임
- 필요한 요소 중 일부는 오픈 소스 자산임. AlphaFold 단백질 구조 데이터베이스 및 다양한 omics 데이터 세트와 같이 광범위하게 액세스할 수 있는 알고리즘은 오픈 소싱 경향의 초기 예시임. 하지만 다른 자산들은 의료 기술 및 데이터 및 분석 회사가 소유하게 되어 더 긴밀한 협업과 파트너십을 강요하게 될 것 임. 많은 분야에서 요구되는 전문 지식의 범위는 단일 제약 회사가 필요한 모든 기능과 기술을 사내에서 개발하는 데 어려움을 겪을 수 있음을 시사함
- 이러한 상황과 R&D 내에서의 빠른 변화 속도를 감안할 때 기업은 데이터, 알고리즘 및 유효성 검사 방법과 같은 구성 요소를 필요에 따라 원활하게 연결할 수 있는 개방형 아키텍처 바이오의료 R&D 혁신 모델이 가장 효과적이라는 것을 알게 될 것임. 이는 기업이 R&D 프로세스의 적시에 적절한 시점에 최상의 방법과 최상의 솔루션을 투자하고 배포할 수 있는 유연성을 제공하는 창의적인 혁신 모델이 될 것임

5. 결론

- 바이오의료 R&D에서 진행 중인 패러다임 변화는 2000년대 초반 학습 및 확인 접근 방식(learn-and-confirm approach)과 비슷한 규모임. 그 전에는 약물 발굴 및 개발 아이디어가 후기 단계 시험까지 체계적으로 우선 순위가 지정되고 검증되지 않아 성공률이 낮았음. 학습 및 확인 모델(learn-and-confirm model)은 프로세스와 고품질 파이프라인을 더욱 엄격하게 만들었음. 하지만 파이프라인 깔때기는 근본적으로 변경되지 않았으며, 깔때기의 후반 단계에서 나타난 제한된 양의 학습만 차기의 앞 단계에 알리는 방식으로 순차적으로 유지되었음
- 현재의 기술 발전은 이제 데이터 피드백 루프에 의해 빠르게 추진되는 덜 우연적이고 보다 결정적인 순환 바이오의료 R&D 가치 사슬을 형성하면서 이러한 접근 방식을 붕괴시키고 있음. 새로운 패러다임은 여전히 진화하고 있으며 최종 형태는 분명하지 않음. 본 보고서에서 설명하는 것은 앞으로 나아갈 수 있는 한 가지 방법에 불과할 수 있음. 그러나 데이터, 자동화, 컴퓨팅 성능 및 AI의 발전과 생물학의 결합이 바이오의료 R&D에 대한 전통적이고 환원주의적인 접근 방식을 개선하여 환자 결과를 개선할 것이라는 것은 분명하며, 준비할 가치가 있는 미래라 할 수 있음

[보고서 III] 바이오 제약 연구에서의 AI 활용

1. 개요

- 최근의 발전에 따라 바이오의약 R&D에서 생산성을 높이고 성공 가능성을 높일 수 있는 역량을 구축할 수 있는 기회가 많아졌음에도 불구하고 여전히 많은 비용과 시간이 소요되고 있음
- 바이오 제약 분야에서 AI가 급속히 성장하고 있는 지금, AI 접근 방식을 과학적 프로세스 변화에 완전히 통합하여 대규모로 가치를 제공할 수 있도록 최대한 활용하는 것이 필요함
- 본 보고서에서는 바이오 제약 기업이 AI 기반 발굴을 활용하여 환자 혜택을 제공하는 방법과 함께, 왜 지금이 주요 파트너십을 구축하고 AI 연구에 대한 투자 조정 및 집중투자를 통해 자체적 역량 구축을 추구해야 하는 시점인지에 대해 설명하게 됨
- 의약품 R&D에서 연구 단계의 목표는 임상 개발로 성공적으로 전환할 가능성이 가장 높은 양질의 후보약물을 가능한 빨리, 많이 생성하는 것임
- 발굴(Discovery) 프로세스는 역사적으로 수렴적이고 단계적인 모든 단계에서 감소가 발생하는 성공-실패의 깔때기 프로세스였음. 이 프로세스는 초기에 테스트한 화합물의 수를 고려할 때 매우 비효율적이라 할 수 있음. 이상적으로는 이 프로세스는 환자에게 효과적인 약물을 제공할 대상과 관련된 테스트용 화합물만 진행해야 할 것임. AI는 가치 사슬의 모든 단계에서 가장 유망한 화합물과 표적을 식별하여 동일한 수의 리드를 달성하기 위해 실험실에서 더 적은 수의 성공적인 실험을 수행하도록 도울 수 있음

2. AI 기반 신약 개발 산업: 아직까지 영향력을 판단할 수는 없는 상황

- AI 기반 약물 발굴 산업은 시장의 신규 진입자, 상당한 자본 투자 및 기술 성숙에 힘입어 지난 10년 동안 크게 성장함. 이러한 AI 기반 기업은 크게 두 가지 범주로 분류됨 (1) 서비스로서 S/W(SaaS)를 포함하여 바이오제약기업에 AI를 지원하는 제공업체, (2) 이러한 서비스와 병행하여 자체 AI 지원 약물 개발 파이프라인을 보유한 AI 제공업체
(Box-3 “지금 AI 기반 약물 발견의 시기인 이유” 참조).

[Box-3] 지금이 AI 기반 약물 발견의 시기인 이유

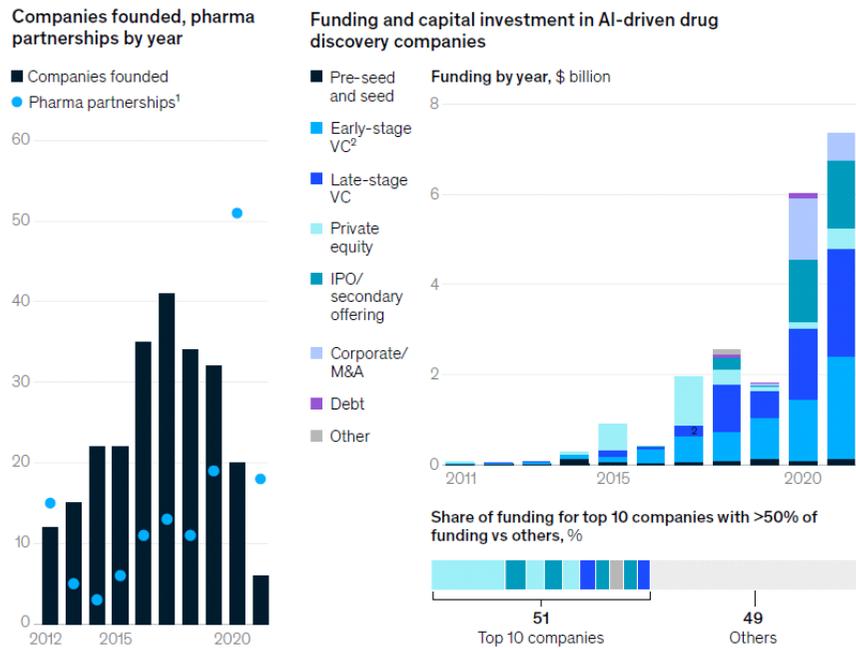
- 기술 발전과 혁신을 위한 규제 개방성이 이제 결합되어 AI 기반 약물 발견을 실용적인 제안으로 만들게 되었음
- 유럽에서는 in silico 및 합성으로부터 파생된 통찰력들에 대한 규제 개방성이 EU 규정 ICH M7 EU(전통적인 분석 기반 접근 방식 대신 정량적 구조 활성 관계(QSAR) 기반 독성 평가를 가능하게 하는 규정)에 의해 촉진되었음. 또한 미국 FDA는 실제사용증거(RWE)를 통합한 대조군과 함께 희귀 질환에 대한 단일군 임상시험을 인정하게 되었음
- 동시에 기술은 두 가지 측면에서 발전하고 있음 (1) 기계 학습(ML)의 산업화 및 확장에 대한 표준화된 접근 방식, 예를 들어 MLOps(ML operations) 및 DataOps(data operations)와 함께 맞춤형 서비스 및 플랫폼, (2) 공공 코드 리포지토리 및 학술문헌을 통해 접근성이 향상된 컴퓨터 비전 및 새로운 분자 설계를 위한 딥 러닝 접근 방식 개발
- 우리의 연구는 AI 기반 약물 발굴 산업에 종사하는 약 270개 기업들을 확인하였으며, 기업의 50% 이상이 미국에 기반을 두고 있으나 서유럽과 동남아시아에서 주요 허브들이 부상하고 있음
- 자체 파이프라인을 보유한 AI 기반 기업의 수는 현 시점에도 여전히 상대적으로 적은 상황(약 15%가 전임상 개발 자산을 보유). 임상 단계(1-2상)에서 새로운 분자 물질(NME)을 가진 기업들은 주로 라이선싱 계약을 통해 확보한 자산을 가지고 있거나 전통적인 기술을 사용하여 개발된 자산을 가지고 있음
- AI 기반 신약 개발의 성장은 기존 바이오 제약 기업들의 관심을 끌었으며 전통적인 바이오 제약기업들과 AI 기반 제약기업들 간의 파트너십이 급격히 증가하였음(그림 2)
- 그러나 높은 가치, 다수의 계약, 막대한 자본 조달(지난 10년간 투자 CAGR 62%)을 가지고 있는 소수의 AI 기반 기업들을 대상으로 파트너십 활동 및 자금 조달이 집중되어 있는 것으로 나타남. 이 영역에 투자된 자본의 절반 이상이 단 10개 회사(모두 미국 또는 영국에 기반을 두고 있음)에 집중되어

있음. 이는 부분적으로 바이오 제약 기업과 투자자가 AI 기반 기업들의 룰테일을 평가하는 데 어려움이 있기 때문으로 파악됨

- 우리는 이 분야에 깊은 관심을 가지고 있는 바이오 제약기업들이 새로운 플레이어가 무엇을 하는지, 가치 사슬을 따라 어디에서 운영하는지, 기술의 독창성, 어떤 기술이 입증 가능한 영향을 미치는지에 대해 결정하는 데 어려움을 겪는 것을 확인하였음

그림 2: 최근 10년간 AI기반 신약발굴 기업 창업 및 투자 현황

Investment in emerging AI-driven discovery players increased for a decade before the recent public-market downturn.



¹Includes companies founded from previous years.
²Venture capital.
 Source: IQVIA Pharma Deals; Pitchbook (data has not been reviewed by PitchBook or IQVIA Pharma Deals analysts)

- 바이오 제약 기업과 AI 기반 신약발굴 기업 간의 파트너십에서 AI 지원 효과를 발휘하려면 두 가지 잠재적 장애물을 극복해야 함
- 첫째, AI 기반 신약발굴 접근 방식(파트너십 포함)은 내부의 일상적 R&D 활동과 거리가 있는 경우가 많음. 이 방식은 실험으로 진행되며 대규모 영향 달성을 위해 바이오 제약기업의 과학 및 운영 프로세스에 고정되지 않음
- 둘째, 내부 R&D 팀 내에서 디지털화된 약물 발견 기능 및 데이터 세트에 대한 투자는 바이오 제약의 엔드-투-엔드 기술 스택 및 기능을 구축하는 대신 파트너 플랫폼 활용 및 IP 강화에 더욱 자주 사용되고 있음

- 이러한 장애물을 극복하면 파트너십이 결실을 맺을 수 있음. AstraZeneca와 BenevolentAI는 오랜 협력을 통해 특발성 폐섬유증에서 다수의 새로운 표적을 식별했으며, 이후 다른 치료 영역(TA)으로 범위를 확장하였음
- Sumitomo Dainippon Pharma는 Exscientia와 협력하여 강박 장애에 대한 DSP-1181을 약물 발견 프로세스에 일반적인 소요 시간의 1/4에 해당하는 기간에(12개월 미만 대 4년 6개월) 식별할 수 있었음

- 마찬가지로 바이오 제약기업 내에서 AI 지원 기능 구축은 어려우며, 혁신 추진을 위한 다기능 팀을 구성하는 것도 어렵고, AI 지원이 상대적으로 고립된 방식으로 구현되는 경우가 많음. AI 지원 접근 방식은 일상적인 연구 활동에 완전히 통합되지 않은 AI 기반 도구를 사용하여 일상적인 과학과 별도로 수행되는 경우가 많음

- 따라서 바이오 제약기업들은 내부 역량 구축과 AI 기반 신약 개발 회사와의 파트너십 간에 균형을 유지해야 함. AI 영역에서 성공적인 바이오 제약 파트너십을 달성하려면 다음과 같은 핵심 이점을 제공해야 함 :
 기술(AI 플랫폼, 알고리즘 및 인프라), 데이터(선별된 세포 이미지, 스크리닝, ADMET 데이터), 인재(바이오제약 인재를 교육하며 AI 파이프라인을 구축할 수 있는 데이터 과학자 및 데이터 엔지니어의 준비된 공급) 및 환자 영향을 극대화하기 위한 매우 구체적인 전략적 의도와 관련된 데이터 보호 보장(예: 특정 TA에서 특정 분자 클래스를 공동 개발)

3. 기업 내부에서 전사적 역량을 구축하는 데 따르는 실질적인 영향

- 바이오 제약기업이 AI 프로세스를 일상 과학에 성공적으로 통합하고 올바른 기술 세트(데이터 과학, 엔지니어링, 소프트웨어 개발, 역학, 발견 과학, 임상

및 디자인)를 갖춘 교차 기능 팀을 구성할 때 우리는 가치사슬에서 상당한 영향력을 관찰하였음

그림 3: 연구 가치사슬에서 AI의 역할

AI has already delivered value across the research value chain.

Examples of AI-driven innovation in biotech and biopharma and observed impact across the research value chain

	Target identification	Target validation	Hit identification	Lead generation /optimization	Preclinical
Examples of AI-driven acceleration	Insights from data sources (internal and from vendors) to generate novel target hypothesis	<i>In silico</i> , phenotypic, cellular models validate targets/ identify biomarkers	Automated image analysis for cellular assays through computer vision technology	Molecular structure and property prediction (eg, protein binding, logP, toxicity) for novel target proteins	Safety issue and DMPK ¹ prediction using internal and public data
	Gene network, biochemical pathway, and cellular-response data integration in target identification	Disease causality determined within patient groups with significant unmet medical need	Molecular property prediction (virtual screening)	Rapid design iteration, across small and large molecules, using eg, Generative Adversarial Networks	Hypothesis-driven dosages for adaptive trials and targeted populations
Examples from industry and observed impact	Biopharma unlocked all-inclusive view of complex indication by attributing disease causality through linkages between genomic data and patient electronic medical records (EMRs)	Biopharma internalized AlphaFold2 and ColabFold to generate 3-D models of almost any known, synthesized protein and protein-protein interactions, reducing access to 3-D structures from 6 months to a few hours	Biotech saw significant acceleration of high-throughput screening (HTS) phase (time to 75% hits detected reduced by 50%) with platform-based "compound prioritization" algorithm	Biopharma leveraged generative machine learning model to expand library/ optimize promising compounds and predict compound efficacy, significantly increasing efficiency of library expansion, with >60,000 new compounds generated	Biopharma utilized predictive algorithms to maximize probability of successful PK ² predictions with 83% of drug development projects progressing to clinic with no PK issues

¹Drug metabolism and pharmacokinetics.
²Pharmacokinetic.

- **가설 생성 기능** : 표적 식별을 위한 지식 그래프를 통해 실제 데이터(RWD), 유전체학 데이터 및 과학 문헌을 결합하여 연구자의 시간이 몇 주간 소요되는 작업이 몇 분 만에 선별된 목록을 만들어내는 실험 생물학 분야의 단순화된 가설 생성 작업

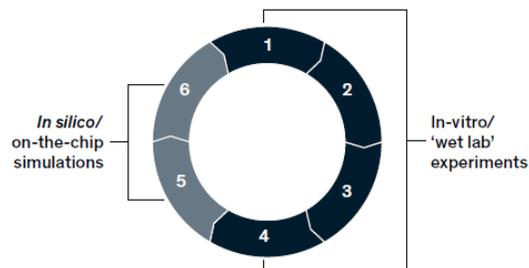
- **거대 분자 구조 추론** : 표적 식별을 위한 단백질 구조 생성(예: 펩타이드 또는 mRNA-백신-항원 생성)까지 시간을 100배 가속
- **컴퓨터 비전 기술** : 기존 접근 방식보다 더 높은 정확도를 가진 최대 10배 가속화된 스크리닝-플레이트-이미지 분석, 타겟 검증 및 히트 식별을 위한 딥러닝 접근 방식(예: 컨볼루션 신경망) 활용
- **In silico 의약 화학** : 히트 식별을 위해 반복 스크리닝 루프에서 분자 특성 예측과 같은 접근 방식을 사용하여 저분자, 고효율 스크리닝에서 30-50% 가속(화합물 무작위 선택의 기존 접근 방식 대비)
- **In silico 화학정보학** : “관찰된 효능” 지표에서 기준 대비 2배 이상 개선, 이전 스크리닝 대비 가능한 in silico 실험 횟수 100배 이상 증가, 약물 전달 최적화 및 리드 최적화를 위한 화합물 설계 시간 단축
- **지식 그래프 기반 가설 생성 및 약물 용도 변경** : 특정 질병 표현형과 관련된 게놈 정보 및 경로를 통해 기존 IND(연구용 신약) 또는 시판 약물에 대한 새로운 적응증을 신속하게 식별하여 R&D 전임상 단계의 일환으로 환자를 위한 새로운 치료법에 대한 시간을 단축
- **유전체학을 활용한 적응증 탐색** : 새로운 작용 메커니즘(MoA)을 추구하기 위한 적응증의 우선순위 지정, 수명주기 관리를 위한 새로운 미개발 적응증 탐색, 성공률이 낮은 프로그램을 조기에 중단하고 임상 시험에서 환자 부담을 줄임으로써 임상 계획 내에서 진행 중인 프로그램의 우선순위를 지정하거나 우선순위를 조정 등

- AI 구현의 영향을 극대화하는 바이오 제약 회사는 MVP(minimum viable product, 최소 기능 제품) 개별 사용 사례를 넘어 연구 시스템을 구축할 수 있음(그림 4)

그림 4: AI 기술을 적용한 연구 시스템 사례

Research systems harness synergies when AI technologies are embedded into 'wet lab' scientific processes.

Parts of a high-throughput screening (HTS) process embedded with AI technology



1. High-throughput screen commenced with diverse compound sets

Scientist selects diverse compound sets (a set of chemical compounds with a wide range of chemical structures) as first high-throughput screen

2. Automated compound selection and transfer

Using HTS machinery, individual compounds are transferred to individual wells of cells under experimental conditions

3. Computer-vision-based hit selection

Cell response to each compound is measured using microscope analysis (eg, through computer vision techniques); promising compounds are labeled "hits"

4. Automated machine learning (ML) model training from screen outcomes

Information from HTS for first few plates is automatically transferred into an ML pipeline, which "learns" how cells respond to each kind of chemical structure

5. Compound library inferencing and prioritization

ML algorithm then scans the remainder of the library compounds and predicts which plates should be prioritized to identify the highest number of hits in the next screen

6. Automated compound selection based on ML recommendations

ML recommendations are automatically queued and used in the next round of HTS. The cycle continues, with the algorithm continuously learning from "real world" outputs. Recommendations trigger scientists to explore new chemical space and begin downstream screening processes more quickly. These recommendations feed into the selection of chemical compounds in step 1

- 연구 시스템은 실험 결과를 향상시키기 위해(단순히 고립된 실제 실험의 준비 단계 대신에) AI를 연구 엔진의 중심에 두어 시너지를 활용함
- 연구 시스템은 AI 알고리즘의 예측 기능과 안정성을 개선하고 실험 설계에 정보를 제공하는 피드백 루프 역할을 수행함(Box-4 “주요 제약 AI R&D

용어 용어집” 참조). 예를 들어 “반복적 스크리닝”: 초기 처리량이 많은 스크리닝의 결과는 기계 학습(ML) 알고리즘을 교육하는 데 사용됨

[Box-4] 주요 제약 AI R&D 용어집

- **인공지능/머신러닝 Artificial intelligence/machine learning :**
종종 같은 의미로 사용되는 AI와 ML은 미묘하게 다른 의미를 갖고 있음. AI의 하위 집합인 머신러닝은 새로운 데이터 추가를 통해 설계상 개선되는 알고리즘을 말하며, AI는 보다 광범위하게 특정 목표를 달성하고자 하는 맥락 하에서 주어진 환경에서 사용 가능한 데이터를 기반으로 문제를 해결할 수 있는 컴퓨터 시스템을 의미함
- **딥 러닝 Deep learning :**
머신러닝의 하위 집합인 딥 러닝은 인공 신경망 구조를 활용하며, 일반적으로 보정하기 위한 매우 많은 수의 매개변수를 포함하여 대규모 데이터 세트에서 패턴을 식별하고 이후에 이 보정된 신경망을 기반으로 예측을 수행함
- **분자특성 예측/반복 스크리닝 Molecular property prediction/iterative screening :**
성공적인 약물 후보가 될 가능성을 예측하는 후보 분자 구조의 패턴을 결정하기 위해 기계 학습 접근 방식을 활용하는 기계 학습과 약물 발견의 교차점에 있는 접근 방식. 일반적으로 "반복 스크리닝" 설정에서 ML 알고리즘은 과거의 스크리닝 실험을 기반으로 교육을 받고 스크리닝 라이브러리의 나머지 화합물 중 히트가 될 가능성이 있는 화합물을 예측함. 추가적인 실험이 수행됨에 따라 알고리즘은 정확도와 예측력을 높게 됨. 알고리즘은 성공적인 예측과 실패한 예측 모두에서 학습함
- **지식 그래프 Knowledge graphs :**
여러 소스의 데이터를 통합하고 일반적으로 동일한 도메인(예: 약물 발견)에서 데이터 간에 링크를 생성하는 데이터베이스 유형. 데이터는 개념 간의 성문화된 관계를 포함하는 개념의 웹인 온톨로지와 통제된 어휘에 따라 구성됨. 지식 그래프는 연결 예측 및 인사이트 생성을 포함하여 여러 데이터 소스에서 연결을 추론하는 데 사용할 수 있음

- DataOps/MLOps :

DataOps(data operations)는 품질을 개선하고 고급 분석의 주기 시간을 줄이기 위해 분석 및 데이터 팀에서 사용하는 자동화된 프로세스 지향 방법론임. MLOps(ML operations)는 기계 학습 및 인공 지능의 맥락에서 IT 작업(예: 생산 소프트웨어 배포 및 패키징)에 적용되는 소프트웨어 엔지니어링 업무에 활용됨

- ML 알고리즘은 대상에 대해 가장 효과적인 기본 화합물 구조를 학습하고 라이브러리의 다른 분자를 테스트 우선순위로 제안할 수 있음. ML 알고리즘이 더 많은 데이터를 수집함에 따라 예측이 더 빠르게 정확해지고 선별된 라이브러리의 상대적인 양에 대해 불균형적으로 많은 수의 “Hits”가 식별됨. 이러한 연구 시스템은 전체 비용을 줄이고, 성공 가능성을 높이고, R&D 프로세스를 가속화하고(환자에게 영향을 미치는 시간 또한 가속화) 특정 사용 사례에 대해 완전히 통합됨

4. 바이오 제약 연구에서 AI를 성공적으로 구현하려면 무엇이 필요할까?

- 디지털 및 데이터 과학 도구와 개념을 구현함으로써 바이오 제약은 현재 포트폴리오의 전체 가치를 포착하고 핵심 기술, 역량 및 IP를 개발하여 미래 연구(예: AI 지원 대형 분자 및 항체 설계)를 추진할 수 있음
- 현재 AI 기반 약물 발견 회사는 이미 비용 효율적인 자체 신약개발 파이프라인을 개발하고 있으므로 기존 기업들이 새로운 기술을 표준 연구 프로세스에 완전히 통합할 수 있는 방법을 찾는 데 도움이 될 것임
- 데이터, 기술 및 인재에 대한 접근성을 제공하고 파트너가 중장기적으로 미래의 경쟁자가 되기 위해 회사의 IP를 악용할 위험이 낮기 때문에 파트너십이 하나의 옵션이 될 수 있지만 유일한 방법은 아님
- 따라서 바이오 제약 기업은 통합되지 않은 사용 사례 또는 파트너십에 대한 투자에서 AI를 일상 연구의 필수적인 부분으로 만드는 방법을 찾는 것이 중요함. 이를 위해 고려해야 할 네 가지 영역은 다음과 같음

(1) 전략 및 설계 기반 로드맵 작성

- 바이오 제약기업은 하향식, C 레벨 전략을 개발하여 AI 지원 발견이 미래

성과의 중요한 원동력이 될 방법을 개발할 수 있음. 중요한 측면은 현재 조직의 문제점이 어디에 있는지, 잠재적인 이익이 무엇인지, AI영역 및 경쟁자가 미래에 어떻게 움직일지를 예상하며 조직이 업계를 주도하고자 하는 위치(경쟁력 뿐만 아니라)를 이해하는 것임

- 이 전략은 구체적이고 시간 제약이 있으며 위험에 처한 가치와 연결되어 있어야 하며 R&D, 연구 및 데이터 과학 책임자를 포함하여 고위 리더 간에 강력한 연계 및 후원이 필요함. 이 전략을 뒷받침하는 것은 충분한 리소스(인재, 데이터 및 인프라 투자)를 현실화하는 데 필요한 역량 구축 및 인재 확보를 지원하는 것이며, 외부 파트너십만 추구하는 경우에는 IP 및 역량 구축에 대한 절충안을 인식하는 것이 필요함
- R&D와 디지털 기능 간의 조정은 균형 잡힌 공동 투자(재정 및 관리 시간)를 보장하고 이니셔티브에서 생성된 영향을 적절하게 공유하는 데 가장 중요함. 또한 AI 기반 약물 발견 접근 방식의 어떤 요소가 파트너십과 내부 구축에 의해 지원되는지 신중하게 고려하는 것이 중요
- 본 보고서에서는 발굴 연구의 어떤 부분을 어떤 순서로 다루어야 하는지 결정하기 위해 디자인 사고 방식을 권장함. 여기에는 연구자에게 병목 현상이 있는 2-3단계가 있을 수 있고 AI가 크게 활용될 수 있는 포괄적 공동 연구 프로세스 연구(예: 중요 세포 분석 또는 리드 최적화를 위한 자동화된 이미지 분석)가 포함됨. 디자인 사고 방식은 기업이 AI, 구현 로드맵, 진행 상황과 영향을 추적하기 위한 성공 지표(예: 대상 식별에서 후보 선택까지의 시간, 대상 식별과 관련된 비용)에서 가장 많은 이점을 얻을 수 있는 영역을 결정하는 데 도움이 될 수 있음
- R&D 및 데이터 과학 리더의 경우 고급 분석 사용 사례에만 초점을 맞춰서는 안되며, 데이터 변환 파이프라인 사용, 대시보드 또는 데이터 플랫폼 및 인프라 구축(예: 지식 그래프) 등 기본 자동화와 같은 응용 프로그램을 사용하여 기존 문제를 해결하는 데 상당한 가치가 있음
- 예를 들어, 생성된 모든 전임상 데이터에 대해 단일 데이터 플랫폼을 구축하면 실험 중복을 방지하고 조직 전체에서 데이터 공유를 향상할 수 있음. 경험적으로 수개월의 가설 생성 시간을 며칠로 단축할 수 있음. 이 결과에는 극적으로 빨라진 속도, 사람들이 더 생산적인 작업에 집중할 수 있는 시간, 분석 품질 향상이 포함됨

(2) 분기별 가치 발표(QVR; quarterly value releases)에 중점을 둔 끊임없는 가치 제공

- R&D, 데이터 과학 및 데이터 엔지니어링이 긴밀하게 협력하고 민첩한 방식으로 사용 사례 전달을 반복하는 것이 중요함. 연구 프로세스에는 구축된 도구 및 시스템의 활용을 보장하기 위해 고려해야 하는 특정 제약 조건 및 작업 방식(예: 실험 방법론의 단계 및 핸드오프)이 자주 포함됨(과학적 프로세스 및 표준 운영 업데이트 외에도). AI 지원 제공을 전체적으로 고려하기 위해 리더는 “high-throughput screening” 에 초점을 맞춘 아래 예시와 같이 주요 빌딩 블록을 정렬할 수 있음
- **청사진 Blueprinting** : 가치 사슬 전반에 걸쳐 사용 사례 목록을 개발하고 영향, 복잡성 및 비즈니스 가치에 따라 우선 순위를 지정함. 이후 가장 필요한 사용 사례를 선택함
- **디지털 및 분석 솔루션 Digital and analytics solutions** : 분자 설명자 (예: SMILES 문자열 형태의 분자 구조)를 원하는 출력 또는 적중과 연결하는 스크리닝 알고리즘을 구축하고 자동화함
- **데이터 연속체 Data continuum** : 재사용 가능한 방식으로 실험 데이터 수집(예: FAIR-데이터 원칙 사용); 장비 및 기존 라이브러리에서 마스터 테이블을 구축
- **기술 능력 Tech capabilities** : 데이터 추출 및 자동 수집을 위한 기술 인프라 및 데이터 아키텍처를 설계하고 구축
- **인재 및 민첩한 운영 모델 Talent and agile operating model** : 도구 및 제공 방법론에 대해 데이터 과학, 데이터 엔지니어링 및 번역가/제품 소유자를 지도하고 협업 환경을 통해 제품을 제공하는 방법을 반복적으로 테스트하고 학습함
- **채택 및 확장(변경 관리 포함) Adoption and scaling (including change management)** : ML 기반 알고리즘을 통합하여 새로운 스크리닝 프로토콜 및 실험 전략을 설계함. 전체 연구 조직(리더부터 실험실 기술자까지)이 기업의 달성 목표와 일상 활동이 어떻게 변화해야 하는지 이해시킴

- 주요 AI 지원 사용 사례가 조정되면 지속적인 영향을 입증할 수 있도록 전달 과정이 고도로 조직화되어야 함. 핵심 요구 사항과 잠재적인 시너지 효과를 식별하고 진행 중인 교차 절단 로드맵의 격차를 식별해야 함. 즉, 조직의 요구 사항에 따라 지속적으로 우선 순위를 재지정하면서 QVR에 집중하기 위해 다년 주기로 영향을 미치는 장기 로드맵에서 출발함. 이 접근 방식을 사용하면 필요에 따라 배치된 중요 임무 자산(mission-critical assets)으로 우선 순위 데이터 수집 및 팀 역량을 동적으로 프런트로딩함으로써 AI 사용 사례 개발을 보다 효율적으로 구축할 수 있음(그림 5)

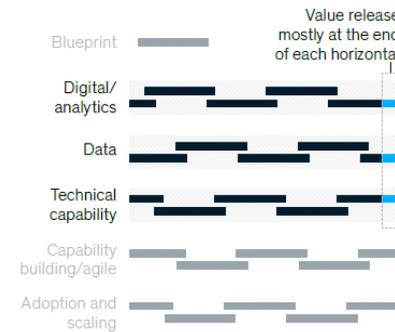
그림 5: 디지털 전달을 위한 6개 빌딩블록(building blocks)

Moving from parallel processes to quarterly value releases touching all six 'building blocks' of an analytics build with links to impact.

Illustration of holistic AI-enablement value delivery focused on quarterly value releases (QVRs) ■ Digital delivery

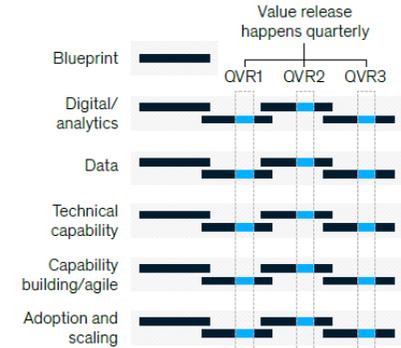
Nonoptimized AI-enablement delivery

Parallel programs of foundational efforts, with minimal linkages between a subset of 6 "building blocks" of analytics builds, where not all blocks may be considered; no explicit link to business/scientific flows and impact



Optimized digital delivery

Program links all horizontals via releases, touching all 6 "building blocks" of an analytics build; links from QVR to impact through scientific/business flows



- 연구의 모든 핵심 디지털 프로세스는 분기별로 점진적으로 제공될 수 있음. 그러나 “가치” 전달의 성격은 다를 수 있음. Moonshot 프로그램(기술 분야에서는 AlphaFold의 출현일 수 있음)에는 바이오 제약에 영향을 미치는 획기적인 발견을 제공하기 위해 장기적인 로드맵과 전용 ML 연구 그룹이 필요함. 이러한 프로그램은 다른 디지털 이니셔티브와 같이 분기마다 AI 제품을 제공하지 않을 수 있지만 인사이드, 보고서 또는 결정은 정기적으로 제공되어야 함

(3) 파트너십을 통한 IP 및 역량 구축, 번역 전문성 개발

- 고유한 기술, 데이터 또는 솔루션 유형에 대한 액세스를 포함하여 특정 영역에서 파트너십은 확실한 이점이 있지만, 이러한 파트너십을 독점적으로 관리하고 새로운 방법이나 솔루션을 일상적인 연구와 분리한다는 것은 약물 발견의 혁신에 필요한 미래 역량이 구축되지 않을 수 있음을 의미함
- 바이오 제약 기업은 파트너십을 통해 제공되는 역량과 사내에서 구축하는 역량에 대해 선택적이고 구체적이어야 함. 마찬가지로 사내 및 외부 인재(특히 우선 순위가 지정된 영역을 지원하기 위해 알고리즘 및 기술 백본을 개발하는 데 연구원과 협력해야 하는 데이터 과학자 및 데이터 엔지니어)에 대한 균형 잡힌 접근 방식이 중요함
- 종종 간과되지만 AI 구현에 중요한 임무를 수행하는 이들은 심층적인 비즈니스, 임상, 과학, AI/ML 및 시스템 아키텍처 이해를 갖춘 “번역가” 또는 “제품 소유자” 임. 이들은 제품 소유권에 대한 사고 방식을 가지고 있으며 분석팀의 모든 요소를 이해하고 동적으로 평가하여 가치에 집중하고 전달에 영향을 미치므로 성공적인 프로젝트 전달을 보장함

(4) MLOps(머신러닝 운영) 및 재사용 가능한 분석 자산을 통한 AI의 산업화

- 바이오 제약 기업이 사내에서 구축하는 기능을 위해서는 올바른 기술 인프라 및 방법론, 특히 DataOps 및 MLOps와 적절한 데이터 아키텍처(예: 그래프 데이터베이스 또는 Data Vault 2.0 기술)와 같은 연구 활동 전반의 확장을 지원할 수 있는 적절한 지원 요소를 갖추는 것이 필수적임. DataOps를 통해 기업은 모델 구축 프로세스를 가속화하여 데이터에서 더 많은 가치를 얻을 수 있음. MLOps에는 AI를 안정적이고 대규모로 제공하기 위한 올바른 팀 운영 모델 및 표준으로 올바른 플랫폼, 도구, 서비스 및 역할을 보장하는 것이 포함됨. AlphaFold, 분자 역학 시뮬레이션, 최적화 모델 및 이미지 인식 워크플로와 같은 컴퓨팅 집약적인 워크플로 처리를 지원하는 기술 아키텍처 인에이블러는 핵심 요구 사항임

- 연구 시스템을 성공적으로 배포하려면 개발 팀은 최종 사용자의 채택을 촉진하기 위해 원활하게 작동하는 여러 상호 관련된 구성 요소(데이터 커넥터 및 파이프라인, 모델, API 및 시각적 인터페이스)를 구축해야 함
- 코드 기반 및 구성 요소의 단편화, 통합 문제로 인한 생산성 감소는 다양한 도메인과 팀에 여러 도구를 배포할 때 발생하는 자연스러운 위험요소임. 개발 시 코딩 표준을 보장하고 팀 전체에서 코딩 접근 방식을 조화시키면 장기적인 생산성과 솔루션 견고성이 향상됨. 또한 조화를 통해 프로젝트 간에 재사용 가능한 구성 요소(데이터 커넥터, 기능 라이브러리, 모델 기반 임베딩)를 공유할 수 있음(예 : 히트 예측을 위한 그래프 신경망 분자 임베딩 및 독성 감소를 위한 리드 최적화)
- 새로운 연구 플랫폼이 복잡해짐에 따라 재사용 가능한 구성 요소의 “자산화”는 점점 더 중요한 개발 생산성(이를 수용하는 팀의 생산성이 두 배)의 원인이 되고 있으며 제품 중심 사고 방식을 가진 전담팀에게 요구되는 중요한 기능이라 할 수 있음

5. 결론

- 오늘날의 질문은 바이오제약 회사가 분석 투자를 개별 프로젝트 및 주요 파트너십을 넘어 연구를 대규모로 혁신하는 방향으로 옮길 것인지 여부임. 특정 과학 및 운영상의 문제에 집중하고 확장 로드맵을 통해 완전히 통합된 연구 시스템에 AI를 구축하는 방향으로 전환하면 바이오 제약 기업이 연구에 AI를 사용하여 실제 비즈니스와 환자에게 미치는 영향을 포착할 수 있음