

The Paradigm Shift in AI Drug Discovery and the Future Convergence Ecosystem

From data-driven target discovery to self-driving laboratories

Target discovery, drug design, clinical validation, and regulation

Kim Kyung-jin, Attorney at Law

Table of Contents

제1부 The Dawn of AI Drug Discovery and the Paradigm Shift	
1.1	R&D Paradigm Shift: Overcoming the Valley of Death
1.2	Artificial Intelligence for Life Sciences: Foundation Models
1.3	Multi-omics Integration and Discovery of Novel Biomarkers
제2부 AI-Driven Drug Design and Molecule Discovery	
2.1	3D Protein Structure Prediction and Structure-Based Drug Design (SBDD)
2.2	Deep Learning-Based Drug Candidate Generation (De Novo Design)
2.3	In Silico ADMET and Toxicity Prediction System
제3부 Next-Generation Modalities and Smart Drug Delivery	
3.1	Biopharmaceuticals and Multi-Target Therapeutics Design (ADC)
3.2	AI Decoding of Natural Product Drug Databases and Hybrid Discovery
3.3	Nanomedicine and Smart Targeted Drug Delivery Systems
제4부 Preclinical and Clinical Validation and the Precision Medicine Revolution	
4.1	Preclinical Innovation: Replacing Animal Testing and Modeling Biology
4.2	Digital Twins and the Modernization of Next-Generation Clinical Trials
4.3	Precision Oncology and Personalized Treatment
제5부 Regulatory Science, Data Ethics, and the Future Convergence Ecosystem	
5.1	Regulatory Frameworks and Quality Governance for AI Drug Discovery
5.2	Data Ethics, Algorithmic Fairness, and the Federated Learning Ecosystem
5.3	The Future of the Pharmaceutical Industry: Self-Driving Labs

PART 1

The Dawn of AI Drug Discovery and the Paradigm Shift

1.1

R&D Paradigm Shift: Overcoming the Valley of Death

In February 2022, the drug candidate ISM001-055 for idiopathic pulmonary fibrosis was administered to a human body for the first time. This disease, where the lungs slowly harden, has no proper cure, and most patients pass away within 3 to 5 years after diagnosis. As lung tissue turns into hard scars, it becomes increasingly difficult to breathe. Because the exact cause is unknown, the word idiopathic is attached to its name just as it sounds. Until then, there were only two drugs in the world that could slow the progression of this disease, and both only slowed the progression to some extent but could not stop it. The company that created this substance is Insilico Medicine, headquartered in Hong Kong. Using the target discovery platform PandaOmics, the research team identified TNIK, an enzyme that regulates the signaling process causing fibrosis, as a new target. They then designed a molecular structure that perfectly fit that enzyme using the chemical molecule generation engine Chemistry42.

This happened in less than 30 months since the project started. If it were a traditional method, this is a phase where finding a target and narrowing down a single candidate alone would have taken 4 to 6 years. This small molecule, which later received the formal name Lentosertib, was thus born on a computer screen and walked into human blood vessels. 3 years later, this substance became the subject of news that once again caught the attention of the entire industry.

Developing new drugs has long been like a tug-of-war across a valley. On this side of the valley, tens of thousands of promising candidates from the laboratory stand in line, and on the other side of the valley, a few pills that are

actually prescribed to patients are placed. Because the number of candidates that fall while crossing in between is overwhelmingly large, the industry has called this section the valley of death. Traditional drug development takes an average of 10 to 15 years from the moment a target is found to receiving approval from regulatory authorities, and the combined cost, including failed candidates, has exceeded 2.6 billion dollars.

This amount includes not only the development cost of a single successful new drug but also the sunk costs of numerous candidates that were tried together and failed. Even so, the substances that cross the valley and reach the market amount to less than 10 percent of the initial candidate pool. The rest collapse in the middle as they lack efficacy, reveal unexpected toxicity, or break down and exit the body differently than expected.

There is a reason why the valley has become so deep. New drugs are first tested on animals like mice or monkeys. However, animal bodies and human bodies receive the same substance differently. An ingredient that safely broke down in a mouse liver can turn into a toxic substance in a human liver, and an effect confirmed in a monkey might not even appear halfway in a human. Even the same human responds differently to the same drug depending on age, genes, and other diseases they already have. This translational research process, which attempts to transfer and apply animal testing results to humans, inherently carries limitations from its birth.

To explain with a metaphor even an elementary school student would understand, it is similar to a situation where a training method that worked on a puppy is used exactly the same on a cat. Although they may look similar on the outside, internal reactions differ by species, so a significant number of candidates that pass animal testing simply collapse in the first clinical trial stage targeting humans. The industry refers to such failures as the interspecies barrier, meaning it worked for mice but did not work for humans.

This cycle of failure also appears clearly in statistics. While the cost to get a single new drug approved has roughly doubled every 9 years, the number of approved new drugs for the same investment amount has rather decreased. Researchers have attached the name Eroom's Law to this trend. It is a word made by reversing the spelling of Moore's Law, which states that semiconductor performance steadily improves. If Moore's Law is a symbol of technological advancement, Eroom's Law is a symbol of the productivity decline that the pharmaceutical industry has not escaped for over half a century.

Experts point out as causes that the scale of clinical trials is growing larger and regulatory thresholds have risen, and that drugs that already work well have occupied most easily developed diseases, making remaining diseases harder to tackle. There are also observations that newly emerging disease targets are only those with complex structures and are difficult to handle. As the situation of increasingly spending more money while results decline has continued for a long time, pharmaceutical companies could not readily jump into high-risk research on rare diseases or intractable tumors.

What has emerged as the key to reverse this trend is data. Information that a human brain cannot fully read has been piling up, including genome and protein information, single-cell data recording the reaction of every single cell, patients electronic health records and imaging data, and thousands of medical papers pouring out every day. For a human to read this information one by one and connect them is not a possible scale to begin with. Past drug development relied heavily on researcher intuition and large-scale screening, that is, randomly testing tens of thousands of chemical compounds to find a substance caught by chance.

Artificial intelligence is shifting the research method itself toward anticipating the causes of diseases and drug reactions in advance by learning this vast data. As searches that randomly scoured for targets changed into predictions based

on data, the proportion of trial and error that dominated the first stage of drug development has noticeably decreased.

The difference becomes even clearer when you place the two pipelines side by side. In a traditional pipeline, 10 to 15 years and over 2 billion dollars of capital go in, it goes through sequential verification where chemists and biologists check by hand step by step, and at the end of it, nine out of ten conclude in failure. In an artificial intelligence pipeline, this period is cut down to less than half, the cost for initial exploration is vastly lowered, and on a structure where multiple stages run simultaneously, generative artificial intelligence continues to refine the candidate itself.

It does not stop at saving time and money, but the approach itself has changed toward filtering out candidates highly likely to fail from the very beginning. Instead of confirming failure later at a high price, the center of gravity has shifted toward checking cheaply and quickly in advance.

The traditional new drug development pipeline follows the sequence of target identification and compound search, optimization, preclinical testing, and clinical phases 1, 2, and 3 lined up in a row. Since each stage must be completely finished before moving to the next stage, if time is delayed at any one stage, all the subsequent stages are delayed together. The pipeline with artificial intelligence greatly disrupts this order. Target identification, compound candidate design, and the work of predicting how the substance will move in the body all run simultaneously on multiple computers.

While generative artificial intelligence creates thousands of molecular structure alternatives at once, another research team pre-simulates the toxicity of the most promising candidates. Work that waited in line for its turn has changed into work that progresses simultaneously in multiple places. This parallel structure is the actual driving force that has cut development time to less than half.

The effects are visible in numbers. Traditionally, finding a new disease target and selecting the first compound to match it took years, but biotechnology companies that adopted artificial intelligence platforms have cut this time to between 12 and 18 months. The number of compounds that actually need to be synthesized for verification has also been greatly reduced from tens of thousands to hundreds. Global management consulting firms predict that if artificial intelligence is introduced into the entire new drug development process, new economic value of 60 billion to 110 billion dollars per year will be created for the entire pharmaceutical industry. Much of this value comes from reducing sunk costs that disappear when candidate substances fail at late stages, and the remainder comes from shortened development periods allowing drugs to be sold longer within the period protected by patents.

Looking at Insilico Medicine alone, it took 13 to 18 months from target identification to preclinical candidate confirmation, and the cost involved was around 2.6 million dollars. With the traditional method, the same period would have taken 2.5 to 4 years and tens of millions of dollars. Looking at just the year 2025, venture capital funding for artificial intelligence-based drug development companies reached 2.7 billion dollars by the third quarter. Capital has jumped on this trend first.

This cost-saving effect has changed the calculation for the entire industry. Previously, it was difficult to find a justification for investing over 10 years and more than 2 billion dollars in rare diseases affecting only thousands of patients. Because the failure rate was so high, few pharmaceutical companies were willing to take such risks for a small number of patients. However, as the initial discovery cost has fallen to the level of millions of dollars and failure can be filtered out at much earlier stages, money is flowing again into rare diseases that had been neglected, pediatric cancers, and research on treatments for antibiotic-resistant bacteria.

This flow is important not just because of statistics or stock prices. Diseases like idiopathic pulmonary fibrosis, which had been pushed down in priority for new drug development due to small patient numbers, are now returning to researchers' desks thanks to lower initial development costs. The wider the bridge that crosses the valley becomes, the greater the possibility that the variety of treatments reaching patients through that bridge will also increase.

The term "generative artificial intelligence" refers to artificial intelligence that does not stop at finding a single correct answer, but instead creates new results that match the given conditions. By the same principle as creating a new picture when asked to draw one or a new sentence when asked to write one, in drug development it creates new molecular structures that match the target. A foundation model is not a model trained on a single specific disease, but rather refers to a foundational model that learns broad data including genes, proteins, compounds, and academic papers and is used together for multiple problems.

Once a model is created, it can be divided and used for multiple tasks from target identification to molecular design and toxicity prediction, making it much more efficient than creating a new model from scratch each time.

The way these two technologies actually design new drugs is easy to understand if explained with a key story. Virtual screening in the past was the work of trying keys one by one from a library of millions of existing compounds to find one that matched the target. In contrast, recent generative artificial intelligence looks at the lock - that is, the three-dimensional structure of the target protein - and then carves out a new key that fits precisely into that hole. Researchers use technologies such as generative adversarial networks, diffusion models, and graph neural networks to pre-visualize in computers the binding affinity and selectivity of drugs, as well as the processes of absorption, metabolism, and elimination in the body.

This method of designing compounds from scratch is called de novo design. New drug candidate discovery has moved beyond the realm of searching through existing substances into the realm of creation, where new substances are built. In the sense that even a molecule never seen anywhere in the world can become a candidate if it fits the target perfectly, the scope of candidates that can be found from the start has itself broadened.

There is an early example showing how fast this design method is. Exscientia, a British artificial intelligence-based drug development company, partnered with Japan's Sumitomo Dainippon Pharma to design DSP-1181, a candidate drug for treating obsessive-compulsive disorder. This substance, designed to target a type of serotonin receptor, was completed less than a year from the start of exploratory research, whereas it would have taken an average of 4.5 years with the traditional method.

In January 2020, DSP-1181 entered its first human clinical trial in Japan. Since this happened just 12 months after computer design began, the pharmaceutical industry was greatly stirred by this news at the time. However, this substance's development was later suspended, and it also serves as an example that later revealed the fact that passing quickly through the beginning of the valley offers no guarantee of safely reaching the other side.

Artificial intelligence that reads structures has also developed alongside this. The AlphaFold series, which predicts the three-dimensional structure of proteins at the atomic level, has discovered hidden binding sites on proteins that were once considered difficult to target, and has greatly increased the speed of molecular docking simulations. Demis Hassabis and John Jumper, who developed AlphaFold, received the 2024 Nobel Prize in Chemistry for this contribution. As the problem of predicting protein structures, which remained a difficult challenge in biology for over half a century, was solved by artificial intelligence, the time it took to identify the target structure - the first step in designing new drugs - shrank from several months to a few days. Tx-LLM, a

therapeutics-specific language model released by Google, trained on molecular structures, target proteins, and clinical literature together in a single model, recording state-of-the-art performance on 22 out of 66 drug development evaluation tasks.

In April 2025, Google released a new open-source model called TxGemma, which advanced this model, and it achieved results that matched or exceeded the previous state-of-the-art levels on 50 out of the same 66 tasks. Evo, a foundation model that handles the entire genome, is also automating the target discovery process by reading single-cell responses and protein-protein interactions together.

Molecules designed inside a computer like this are now made into actual substances with almost no human intervention. In automated laboratories where robot arms move reagents and control reaction conditions, hundreds of synthesis pathways suggested by artificial intelligence are tried day and night. The results of measuring the effects of the synthesized substances are immediately reflected in the next design, forming a closed-loop system where design, synthesis, and measurement are repeated without human intervention. When standardized robots take over experiments, the variation in manual skills that appeared differently for each person decreases, which also reduces the problem of conflicting results when experiments are repeated under the same conditions.

The scene of researchers staying up all night holding reagent bottles themselves is changing into a scene where people only set the goals and machines run hundreds of experiments overnight on their behalf. These autonomous laboratories are expanding the scope of parallel processing outside the computer screen and onto the actual laboratory bench. Since experiments run continuously even while people are asleep, the actual operating time of the laboratory, which used to be roughly eight hours a day, has significantly increased.

However, for a blueprint that looks perfect on a computer screen to become an actual building, a long construction period is still needed. Even virtual compounds refined by in silico simulations can show unexpected off-target reactions, hepatotoxicity, or cardiotoxicity once they enter a living body, so they must undergo preclinical validation using in vitro cell experiments, animal experiments, and organoids.

Even though artificial intelligence has reduced the time it takes to find candidate substances from years to months, Phase 1, Phase 2, and Phase 3 clinical trials that confirm safety and efficacy in the human body cannot be skipped because patients' lives are at stake. This verification takes several months at the shortest and several years at the longest. This means that while the width of the valley can be narrowed, the valley itself cannot be eliminated.

Preclinical validation is a process of checking five main things. How the drug is absorbed in the body, which organs it spreads to, how it is broken down in the liver, how it exits the body, and whether it causes unexpected toxicity are collectively called absorption, distribution, metabolism, excretion, and toxicity. In the past, these five things had to be checked one by one through animal experiments and cell experiments, but now artificial intelligence models offer close predictions first just by looking at the molecular structure.

Although these predictions cannot completely replace actual experiments, being able to filter out candidates with a low chance of success before they even reach the laboratory bench saves both the time researchers spend and the number of laboratory animals mobilized.

The ways candidate substances fail inside the valley are also varied. Some substances stick well to their targets in cells inside a petri dish but scatter before reaching the desired tissues in the human body. Some substances are broken down in the liver so quickly that they exit the body before their medicinal effects can appear. Other substances intend to touch only one target but end up touching other proteins with similar structures, causing

unexpected side effects.

Artificial intelligence has separate prediction models corresponding to each of these failure types, allowing researchers to estimate from the initial design stage at which step a candidate substance is likely to fail. Weaknesses that used to take weeks to uncover in the laboratory are now increasingly being noticed first during the stage of drawing the molecular structure.

An example that painfully confirmed this fact emerged in 2023. Ulotaront, a candidate substance for treating schizophrenia developed by Sumitomo Pharma together with Otsuka Pharmaceutical, is a substance discovered using PsychoGenics's artificial intelligence-based behavior analysis platform. Unlike existing schizophrenia drugs, it gathered expectations that it would reduce side effects by targeting other neurotransmitter receptors instead of dopamine receptors. However, in July 2023, in two Phase 3 clinical trials named Diamond 1 and Diamond 2, Ulotaront failed to improve symptoms significantly better than a placebo.

There were no problems with safety itself, but it failed at the final hurdle where it actually had to prove its efficacy. Analyses also suggested that interpreting the results became more difficult because it overlapped with the COVID-19 pandemic, causing even the symptoms of the patient group receiving the placebo to improve much more than in previous years. This failure shows that no matter how precisely artificial intelligence predicts, it cannot perfectly control variables in the clinical field, such as individual patient reactions and the placebo effect.

To reduce these uncertainties that happen in the clinical stage, the solutions the industry is paying attention to are digital twins and adaptive clinical trial designs. A digital twin is a technology that creates a virtual twin by moving a patient's genetic and pathological characteristics exactly as they are into a computer. Before administering actual drugs, researchers simulate how tumor heterogeneity or drug resistance changes by applying various doses and

treatment scenarios to this virtual patient in advance.

This approach carefully divides patients, refining patient stratification to first test drugs on those predicted to respond well. More attempts are also using virtual control groups built on real-world data to reduce the size, length, and ethical burden of clinical trials. Decentralized clinical trials, where people can participate without visiting hospitals, are also growing along with this trend.

A common reason for clinical trial delays is failing to gather participants on time. Statistics even show that more than 80 percent of all clinical trials fall behind schedule because of delayed patient recruitment. Artificial intelligence is reducing this bottleneck by scanning electronic health records and genomic data to find potential participants who fit the conditions in advance. It also suggests criteria that make recruitment easier right from the trial design stage.

Boosted by this demand, the market size for artificial intelligence specialized in clinical trials is expected to grow from 2.8 billion dollars in 2024 to more than 54 billion dollars in 2032. Gathering people onto the bridge on time is just as important as widening the bridge across the valley, and this also determines the speed of new drug development.

And in 2025, the first case of actually crossing the valley emerged. The Phase 2a clinical trial results for ISM001-055, which entered Phase 1 in 2022 and has now been named Lentosertib, were published in the international journal Nature Medicine on June 3 of that year. The fact that the results were published in a peer-reviewed academic journal, rather than a corporate press release, means that it passed verification by independent experts. This trial, conducted on 71 patients with idiopathic pulmonary fibrosis for 12 weeks, combined a double-blind method where neither the doctor nor the patient knew who received the real drug, with a placebo-controlled method placing a fake drug as a comparison group. Lentosertib was safe at all doses, and lung function improved clearly as the dose increased.

The patient group receiving 60 milligrams a day saw their lung capacity indicator increase by an average of 98.4 milliliters, while the patient group receiving the placebo actually saw it decrease by 20.3 milliliters. These results were also presented at the American Thoracic Society conference the same year. This is the industry's first publicly verified case where targets and molecules born inside a computer led to visually confirmed effects inside the bodies of actual patients.

This is not a story just about Lentosertib. As of early 2026, there are 173 artificial intelligence-based new drug candidates that have entered the clinical stage. Compared to 24 at the end of 2023, this has increased more than seven times in less than 3 years. According to various industry analyses, substances found by artificial intelligence reach a Phase 1 passage rate of 80 percent to 90 percent, far exceeding the 40 percent to 65 percent of traditional methods. The purpose of Phase 1 is not efficacy but confirming safety. Because artificial intelligence filters out structures expected to be toxic right from the design stage and simulates the absorption and excretion processes in the body in advance, it seems that only candidates that easily cross the safety threshold enter the clinical trial stage in the first place.

However, the gap narrows when moving to Phase 2. So far, four out of ten artificial intelligence-discovered substances that completed Phase 2 have succeeded, recording around 40 percent. This is almost the same level as the industry average of 30 percent to 40 percent. From Phase 2 onward, variables that are difficult to control in the design stage, such as each person's different genetic background, lifestyle habits, and accompanying diseases, seem to take on a larger role than the precision of the drug design. As of the summer of 2026, there is not yet an artificial intelligence-designed new drug that has received final approval from regulatory authorities.

This seems to mean that while the entrance to the valley has noticeably widened, the cliff at the edge of the valley remains steep.

Systems are also pushing this trend from behind. In December 2022, the United States Food and Drug Administration passed the FDA Modernization Act 2.0, eliminating the decades-old mandatory clause that new drug approval applications must include animal testing. In April 2025, starting with monoclonal antibodies, it laid out a specific roadmap to replace animal testing with organoids and artificial intelligence-based simulations. The subsequent FDA Modernization Act 3.0 passed the Senate in December 2025 and even passed the House of Representatives on July 20, 2026.

Because the structure requires the Senate to vote once more on the bill revised by the House of Representatives before it goes to the President, it has not yet been signed as of August 2026. When this law goes into effect, the phrase animal testing embedded throughout new drug approval regulations will disappear. Non-animal tests, such as computer simulations and organ chips that recreate the functions of organs like the lungs or liver on a fingernail-sized device made of human cells, will replace that phrase.

Large pharmaceutical companies have also rushed to join this trend. Isomorphic Labs, a subsidiary of Google DeepMind that created AlphaFold, signed multi-billion dollar new drug co-development contracts with Eli Lilly and Novartis, respectively, and is jointly conducting molecular design for various disease targets. Sanofi reorganized its entire organizational structure with the goal of integrating artificial intelligence into its entire research and development decision-making process.

Around the same time, other large pharmaceutical companies, including Novo Nordisk and AstraZeneca, also embarked on similar reorganizations by creating their own artificial intelligence organizations or partnering with specialized companies. Aside from specialized artificial intelligence drug development companies independently designing new drugs, the trend of existing large pharmaceutical companies transplanting the same tools into their own research and development organizations is also growing.

The capital market's response is also clear. Insilico Medicine went public on the Hong Kong Stock Exchange in late 2025, raising 2.277 billion Hong Kong dollars, or about 292 million dollars, recording the largest initial public offering among Hong Kong biotech companies that year. The public subscription competition ratio exceeded 1,427 to 1, and 15 institutions, including Eli Lilly, Tencent, and Temasek, were listed as cornerstone investors. Similar movements are also accelerating within Korea. In February 2026, the artificial intelligence-based new drug development company Galux attracted a 42 billion won Series B investment involving Yuanta Investment, Korea Development Bank, Korea Investment & Securities, and Mirae Asset Securities.

Companies like Brize Bio, First Biotherapeutics, and Elysigen are raising investments exceeding 3 billion won in succession and laying the foundation for going public. OncoVigil is designing small-molecule targeted therapeutics using its own generative AI platform, and AronTier is partnering with Celltrion to discover new protein structure-based candidate substances. Although most of these companies are still in the early stages of clinical trials, looking at the scale of investment and the profiles of participating institutions, it is read as a signal that domestic capital's perspective on this field has shifted from before.

The industry views 2026 and 2027 as the real test. Until now it was a time to confirm speed and early pass rates, but the coming years ahead will be a time for the leading pipelines to receive their report cards in front of phase 3 clinical trials, the final gateway. Depending on this report card, how pharmaceutical companies will allocate R&D funds over the next ten years is expected to differ greatly. This is likewise a test that comes to Korea's new companies.

Whether Lentosertive can reproduce the same effect in the next phase 3 clinical trial, and whether that result leads to final approval by regulatory authorities, is expected to be a measure for gauging the trust of AI-designed drug development as a whole. The valley of death has not disappeared, but the bridge crossing over it has just finished its first vehicle passage.

The Korean government is also preparing a support system to match this trend. The Ministry of Health and Welfare and the Ministry of Science and ICT, with the Korea Pharmaceutical and Biotechnology Industry Association as the project executive agency, are carrying out the K-Melody project from April 2024 to December 2028 with 34.8 billion won. This project, modeled after Europe's Melody Project, does not gather drug candidate substance data accumulated separately by each company in one place, but instead uses a federated learning method that keeps data in each institution as is while moving only AI models around and training them.

From the pharmaceutical company's perspective, records of failed experiments are sensitive information they do not want leaking to competitors, and this method has opened the way to learn together without sending data outside. In an interview, Director Kim Hwa-jong, who leads the project, said that AI must accumulate more cases of failure than success in order to filter out the next failures in advance. This means that failure is also data.

The Ministry of Food and Drug Safety also released a guideline for AI utilization in drug development in June 2024, providing direction on how to handle materials created by AI in the new drug review process. Daewoong Pharmaceutical entered the K-AI new drug development preclinical and clinical model development project as a joint research institution, and decided to provide nonclinical data accumulated through research on anticancer and metabolic disease treatments to AI model learning over four years.

The company has also separately created and is operating a platform called Daisy for finding drug candidate substances based on its own database containing 800 million types of compound molecular models. Not only emerging venture companies but also pharmaceutical companies with long histories have begun to hold the same tools.

Yet it is natural to ask why there is still no case of AI-designed drugs that has reached regulatory approval. Even DSP-1181, which entered clinical trials first, was first administered to humans only in 2020, and Lentosertive only began its first clinical trial in 2022. If we directly apply the time period required to go through phase 1 to 3 clinical trials and regulatory review, even candidate substances currently ahead would need to reach 2027 at the earliest or 2028 at the latest for results to come out.

It is more likely that not because AI-designed drugs are failing at an unusually high rate, but because sufficient time has not yet passed for it to lead to approval. No matter how quickly you run a calculator, you cannot speed up the time it takes for a patient's body to respond to a drug by several years. The phase 3 clinical trial results that will emerge in the coming years are expected to actually confirm whether this judgment is correct.

1.2

Artificial Intelligence for Life Sciences: Foundation Models

In February 2025, before a computer at Arc Institute, a nonprofit research institute in California, researchers were watching vast genetic sequences from humans and rice, E. coli and archaea flowing continuously into their model without pause. What crossed the screen was not a familiar gene name but sequences made up of only four letters: adenine, guanine, cytosine, and thymine. Yet within them were mixed the rules of all kinds of life phenomena, from human heart disease to antibiotic resistance in microorganisms.

The program they were building was not a calculator designed for a single disease or a single organism. Until just a few years earlier, artificial intelligence used in biology research labs had remained a narrow tool that fixated on one gene or one target protein. For this to attempt to provide answers with a single model regardless of species was itself a departure from convention. Whether it was a human genetic disease, a plant's drought resistance, or enzyme design for industrial microorganisms, the goal was to have the same single model provide answers even when the questions being asked were different.

For those making medicines, genomes had been used only as reference materials, and no one readily thought of them as objects that could be dealt with as a whole, calculated like predicting the next sentence. A few weeks later, the name of this model released to the world was Evo 2, and its emergence clearly changed the direction that artificial intelligence had been taking in the pharmaceutical and life science fields.

When faced with a new order, a chef who cooks only kimchi stew well and a chef who has thoroughly mastered the basics - blade work, heat control, and seasoning balance - get different results. The first chef can cook kimchi stew

amazingly but gives up when ordered to make pasta. The second chef, having mastered the basics, can immediately create new dishes just by looking at ingredients in the refrigerator. Even when unfamiliar ingredients come in, the second chef does not panic and applies basic skills to produce believable flavors. The artificial intelligence that the term "foundation model" refers to is closer to the second type of chef. Previous artificial intelligence programs were like the first chef. When trained to score the binding strength between a specific protein and a specific compound, they could do only that one calculation.

If you asked the same program whether a compound would cause toxicity in the liver, it could not provide an answer, and you would have to train a completely different program from the start. A foundation model first learns the basic grammar of life phenomena from vast data, and then applies its capabilities as needed for whatever new question is posed - whether predicting targets (proteins in the body where drugs should attach), designing new drug candidates, or estimating clinical trial results. Because of this difference, pharmaceutical companies are now changing their research organizations themselves, moving toward having multiple departments share one foundation model with basic skills rather than writing separate new programs for each department and each task.

For this principle to gain strength in new drug development, there was one wall that had to be overcome. Electronic health records from hospitals, genome databases, chemical compound libraries from pharmaceutical companies, and research papers scattered across academic journals had for a long time been locked under different servers, different formats, and different ownership. Genomic data on lung cancer patients compiled by a university hospital circulated only within that hospital, and toxicity data on compounds confirmed through experiments by a pharmaceutical company did not leave that company.

Adverse event reports compiled by regulatory agencies during the drug approval review process also remained mostly in internal filing cabinets of institutions, except for a few publicly available databases. When data was divided this way, even the most sophisticated algorithm could only learn from the limited information in the narrow warehouse in front of it. Starting around 2023, the pharmaceutical industry, academia, and big tech companies simultaneously began work to combine these divided repositories into one. When disease information, patient genomes, and clinical data are put together in one model and learned, the model acquires general-purpose biological reasoning ability - looking at just one genetic mutation in a lung cancer patient, it recalls at once different cancer types, different drugs, and different clinical trial results that showed similar patterns.

This general-purpose reasoning ability is precisely the decisive difference that separates the previous generation of programs, which calculated only one disease or one target, from foundation models.

There are multiple layers to the data that look into life phenomena. The genome is like the complete blueprint for designing a body, changing little from birth, and the transcriptome shows which genes from that blueprint are actually being actively used at this very moment. The proteome shows the actual parts created according to the unfolded blueprint - in other words, the list of proteins - and the metabolome shows the final products created as those parts function. Just as you cannot know which floors of a building have lights on by looking only at the blueprint, you cannot fully know what is happening right now inside a specific cell or patient with only genome information.

The reason foundation models learn the genome, transcriptome, proteome, and metabolome together is that you must see the blueprint, the lighting condition, the parts list, and the waste products all at once to properly diagnose the state of the entire building. When the patient's medical records

and imaging data are added to this, the model is able to also infer together how changes at the molecular level connect to actual symptoms.

The insight that delivered the decisive blow to this integration was the discovery that protein sequences and molecular structures could be read in exactly the same way as human sentences. Proteins are structures where twenty types of amino acids are linked like a chain, and this is not fundamentally different from how twenty-six alphabets line up to form words and sentences. Molecular structures that chemists had long drawn only as pictures could also be written by converting atoms and bonds into a single symbolic sequence. The transformer, a neural network structure originally designed to translate human sentences and predict the next word, worked exactly the same way even when fed amino acid sequences or molecular symbol sequences instead of sentences. As the model repeated training to guess which amino acid or atom would come next, it began naturally to learn the grammar of life phenomena, and on its own came to recognize chemical properties like how proteins fold or their reactivity, without researchers having to teach the rules separately.

Models that learned this grammar were able to invent new protein sequences that did not exist before, or design molecules from scratch that fit perfectly into the hollow binding sites of target proteins. As the power of generative artificial intelligence, which used to handle text, transferred directly onto the languages of biology and chemistry, drug development transitioned from selecting among already existing candidates to building candidates that never existed in the world.

There is a certain logic to the order in which foundation models are created. First, they go through a pre-training stage where they learn the rules of biological phenomena themselves without a specific purpose, using vast amounts of data. At this stage, the model is like a student learning the basics of language, mathematics, and science across elementary and middle school. It

has not yet decided on a specific career path but builds a foundation to tackle any problem that comes its way. Once the model, having built this foundation, goes through a fine-tuning stage where it is retrained for a specific task, it finally gains specialized skills for concrete missions like designing drug candidates or predicting clinical trial results.

TxGemma is an example that followed this exact order. It was created by adding fine-tuning with therapeutic data on top of the foundation laid by the pre-training of Google's general-purpose language model, Gemma 2. The fact that a broader and stronger foundation leads to a greater effect of the fine-tuning built upon it shows once again why data integration is the key factor determining the success or failure of a foundation model.

The structure of a foundation model built this way resembles a three-story building. On the bottom floor, compound libraries, genomes, transcriptomes (data showing how much genes are actually being read inside cells), and millions of clinical documents are stacked like foundation materials. On the middle floor, which refines these materials and sets up the framework, there are large language models that understand sentences and foundation models that specialize in reading protein sequences and molecular structures. The framework refined on the middle floor goes up to the top floor and is completed into actual results: generating drug candidates, prioritizing targets, and optimizing clinical trial designs.

For example, if the middle floor model reads the genomes and related papers of lung cancer patients stacked on the bottom floor and singles out one protein that frequently stands out in lung cancer, the top floor immediately draws up a few drug candidates that bind to that protein. If the materials on the bottom floor are poor, there is nothing to build no matter how sophisticated the middle floor is. And if the middle floor collapses, the power cannot be transmitted to the top floor no matter how good the materials stacked on the bottom floor are. The fact that all three floors must be strong at

the same time to produce useful results in the field of drug development is being proven one by one by the foundation models pouring out between 2025 and 2026.

An example that clearly showed the power connecting the middle and top floors is TxLLM, a large language model specialized for therapeutics released by Google. TxLLM is the result of pushing different formats of data - molecular structures, target proteins, cell lines, and clinical trial texts - together into a single model and training it. Although molecules are expressed as atomic symbols, proteins as amino acid sequences, and clinical trial records as sentences written by humans, TxLLM melted all of these into a single common representation space, making it possible to compare and connect them with one another. The research team tested TxLLM's skills on 66 evaluation tasks spanning the entire drug development process, from target discovery to predicting clinical trial results.

The evaluation items included a mix of problems, from assessing the probability that a specific compound would survive to phase 3 clinical trials, to identifying interactions that might occur when two drugs are administered together. Among these, TxLLM recorded better scores than any specialized model available up to that point in 22 tasks. The fact that a single model performed well on two widely separated tasks - molecular design and clinical prediction - simultaneously was accepted as proof that foundation models could walk out of research papers and be used in actual development fields.

The difficulty that drug development teams actually face is not a shortage of target candidates, but having too many of them. The human genome contains more than 20,000 genes that produce proteins alone, and among them, candidates worthy of becoming drug targets pour out in the thousands when searching through papers, patents, and clinical records. Because research funding and personnel are limited, pharmaceutical companies have no choice but to select a few dozen, or at most a few hundred, of these and delve into

them first. In the past, this selection process relied on the intuition of a few experienced scientists and literature searches. As a result, the phenomenon of research flocking only to already popular targets was repeated.

Models like TxLLM or TxGemma scan through targets, diseases, and the success and failure records of existing drugs all at once to score the probability of a specific target succeeding up to the clinical stage. Therefore, they can bring forward candidates that humans might easily miss, along with the evidence. Unlike the past, when the work of setting target priorities took anywhere from a few weeks to several months, the first draft of the ranking table comes out within a few days now that foundation models are used.

The possibility launched by TxLLM went one step further when Google released TxGemma in March 2025. TxGemma is a model retrained with more than 7 million therapeutics-related training cases, using Google's lightweight language model Gemma 2 as its framework. It was released in three sizes: 2 billion, 9 billion, and 27 billion parameters. Parameters are like internal knobs that the model adjusts its values with during the learning process. The larger the number, the more sophisticated judgments generally become possible. An interactively operating TxGemma-Chat version was also released separately from the prediction-only model. Thus, a method where researchers ask questions as if chatting and the model answers by presenting evidence has also become possible.

The largest version with 27 billion parameters performed equally or better than TxLLM on 64 out of 66 tasks, and on 45 of those tasks it broke the previous records. The decisive difference between TxGemma and TxLLM is openness. Unlike TxLLM, which remained a Google internal research result, TxGemma fully opened its model weights through Hugging Face and Google Cloud's Vertex AI, opening the door for university research laboratories and bioventures without sufficient funding to use foundation models of the same level.

The attempt to read an entire genome as a single language reached its peak with Evo 2. This model, created by Arc Institute together with Nvidia, Stanford University, University of California San Francisco campus, and University of California Berkeley campus, has 40 billion parameters and can read sequences of up to 1 million base pairs (the chemical letter units that make up genes) at a time. The data used for training was more than 9 trillion bases obtained from over 100,000 species of organisms, without distinction between humans, plants, bacteria, and archaea. Evo 2 was first released in February 2025 in the form of a preprint paper, and was later formally published in the international journal Nature after verification.

The model demonstrated that it can accomplish everything from identifying genetic mutations that cause disease in humans in advance to designing new microbial enzymes, and even creating entirely new genomic sequences that did not previously exist from scratch. Arc Institute explained that this model can be used not only in human genetic disease research but also in agricultural breeding and industrial microorganism development, and in the search for new materials. Because a single model simultaneously handles different tasks of diagnosis and design, Evo 2 is cited as a symbolic case demonstrating what kind of power the integration of genomic data can bring to new drug development.

Google DeepMind also released a model in a similar direction under the name AlphaGenome on June 25, 2025. AlphaGenome receives as input a DNA sequence of 1 million base pairs in length and predicts thousands of regulatory signals all at once, including gene expression, chromatin accessibility (a value indicating how easily DNA can be opened and read), histone modifications, and fusion positions. The task of determining whether mutations occurring in non-coding regions - regions of the human genome that do not directly produce proteins - actually cause disease has long been a difficult problem that has troubled geneticists, but AlphaGenome achieved superior performance to existing models on 22 out of 24 evaluation items and 24 out of

26 mutation prediction items.

However, this model was not designed to diagnose disease by inputting individual genomes as a whole, and it was also revealed that it still has limitations in accuracy when reading the effects that regulatory sequences more distant than 100,000 base pairs have on genes. In January 2026, Google DeepMind opened the source code and weights of this model for free for non-profit research purposes only, and as a result, a common foundation was established that allowed genetics laboratories around the world to compare regulatory mutations of different diseases using the same tool.

As patient genome sequencing results that had been scattered across different hospitals came together again on this common model, clues began to emerge even for problems that had no answers using only a single hospital's data, such as diagnosing rare diseases.

On the protein side, ESM3, developed by the American startup EvolutionaryScale, achieved a similar leap. While the widely used AlphaFold family of models placed weight on prediction - identifying three-dimensional structures from given amino acid sequences - ESM3 is a model that has taken a step further beyond prediction into design in that it can create new sequences matching the conditions even when the conditions are given in reverse on the structure or function side. ESM3 learns together the amino acid sequence and three-dimensional structure, and what function the protein actually performs, and has acquired the ability to complete the other two by itself whenever just one of the three is given as a condition.

The model's capability was verified when it was published as a formal paper in the international journal Science in January 2025. The researchers told ESM3 only the conditions that a fluorescent protein should have and had it create the remaining sequence anew, and the protein created this way had a sequence distinctly different from any fluorescent protein found in nature, yet shone brightly. The researchers evaluated this result, saying that the model

had reached through calculation alone what nature would have reached through 500 million years of evolution. The weights and code of ESM3 are open through GitHub for academic research purposes only, and a free-to-use tier has been provided in the application programming interface for general users, so an environment has been created where even graduate students who were only reading papers could directly design and test new proteins.

However, using it commercially on a large scale requires a separate contract, so the question of how much to lower the barrier to access remains an issue still under discussion.

The attempt to mimic the responses of individual cells as a whole left clear milestones in the Virtual Cell Challenge opened by Arc Institute in the summer of 2025. In this competition to identify the model that predicts most accurately the responses that appear when genetic modification or drug stimulation is applied to cells, more than 5,000 researchers from 114 countries registered and more than 1,200 teams submitted results. This competition with a first prize of \$100,000 was conducted based on a data repository called the Virtual Cell Atlas publicly opened together by Arc Institute, and this repository contains organized observation and stimulus response data from more than 600 million cells from multiple laboratories.

This data includes a large-scale drug response dataset provided by the biotech company Tahoe Therapeutics, as well as single-cell data that the Arc Institute automatically gathered by having an artificial intelligence agent search through published papers. The model named STATE, which the Arc Institute released after the competition, calculates in advance without experiments how gene expression will change when a specific stimulus is given to a specific cell line. This competition confirmed through its results that general predictive power at the cellular level only arises when data from multiple laboratories, multiple cell lines, and multiple stimulus conditions are gathered in one place, instead of stopping at data from one disease or one patient.

The Arc Institute announced that it will hold this competition again in 2026. As scattered cell experiment data gathered in one place, clues began to emerge for questions that a single laboratory could not answer on its own.

Entering 2026, this trend grew in size beyond the competition of individual models to the level of reorganizing the infrastructure of the entire industry. OpenAI released GPT-Rosalind, a model dedicated to life science research, on April 16, 2026. Named after Rosalind Franklin, the scientist who left a crucial clue in revealing the DNA double helix structure, this model had its performance revised in June of that year, reducing its token (the minimum unit for processing sentences) consumption by 31 percent compared to the previous version while improving its medicinal chemistry and genome analysis skills. However, this model was not released in a form that anyone could use right away, and access rights were first given only to verified institutions like Amgen, Moderna, Thermo Fisher Scientific, and the Allen Institute.

Given that Amgen and Moderna, which put their names on this list, are large pharmaceutical companies that have already accumulated vast clinical data on their own, this shows the possibility that OpenAI's model will take root by sitting on top of existing data instead of pushing it out. Along with this, OpenAI also started a separate business called Rosalind Biodefense to prevent situations where life science technology is used for wrong purposes. The very fact that they jumped into a dedicated life science model with their company name on the line shows that these foundation models are no longer the domain of academia and a few biotechs alone.

Amazon Web Services (AWS) also introduced Amazon Bio Discovery at a life science symposium on April 15 of the same year. This service took the form of an artificial intelligence assistant that gathers over 40 types of biological foundation models on one screen, and when a researcher makes a request in natural language, it chooses the appropriate model and even creates an experimental design. Amazon Web Services stated that Twist Bioscience and

Ginkgo Bioworks had already put their names on the list of collaborating laboratories in charge of synthesis and verification, and A-Alpha Bio is also scheduled to join soon. This service also has a function that changes it into a single work sequence made up of multiple models and analysis steps when a researcher writes down an experimental procedure in natural language, so even a biologist who does not know how to write code can build a complex pipeline by hand.

Memorial Sloan Kettering Cancer Center designed 300,000 types of antibody candidate molecules with this service and then sent 100,000 of them to Twist Bioscience to complete actual synthesis and verification, reducing work that used to take months to just a few weeks. Nvidia had already added MolMIM, a generative model that handles small molecules, to BioNeMo in April 2025, and in January 2026, it newly placed an RNA structure prediction model and a synthesis possibility verification model on the same platform. On the 12th of that month, they jumped into the work of connecting foundation models to the laboratory site outside the computer by announcing a collaboration with Eli Lilly to build a joint innovation laboratory that solves the difficult problems of new drug discovery together, alongside a collaboration with Thermo Fisher Scientific to graft artificial intelligence onto laboratory equipment.

This cyclical structure that immediately connects prediction and verification shows that foundation models have taken root inside actual new drug pipelines, going beyond the performance indicators in papers.

Training such models from the beginning requires an amount of calculation that runs thousands of graphics processing units for weeks, and the electricity bills and equipment costs alone easily exceed the annual budget of an average biotech venture. As a result, there is a clear trend where the work of building a new model from the bottom up is concentrated in a few companies with their own data centers, like Nvidia, Google, Amazon, and OpenAI, and some laboratories that receive government support.

Biotechs or university laboratories with a small scale are increasingly choosing to download the model weights released by these companies and fine-tune them with their own special data. The fact that TxGemma, ESM3, and AlphaGenome released their weights can also be seen as an attempt to reduce this gap even a little.

The foundation models that have come out so far each have their own areas they are good at. Evo 2 and AlphaGenome read genome sequences well, ESM3 handles protein structures well, and TxGemma weaves text and clinical data well. The work of integrating into a single model that reads a single patient's genome, cell responses, medical records, and imaging data all together is still in progress. A method that gathers multiple models on one screen and lets you choose and use them whenever necessary, like Amazon Bio Discovery, can be seen as an intermediate step leading to complete integration.

The fact that a standard to connect models released by different companies and different laboratories into one has not yet been prepared is a noticeable task that this field faces. Nevertheless, looking at the trend over just two or three years where models that used to read only a single genome have rapidly broadened in a direction that encompasses cell responses and clinical records, we can guess that the next stage of integration is not too far away.

The value of a foundation model is determined more by the breadth of the data it has consumed and the speed at which that data is verified through experiments than by the algorithm code itself. Even if they use the same transformer architecture, a model fed with the genomes of 100,000 biological species and a model fed with only one company's compound data show completely different capabilities. Therefore, the question is now shifting from which algorithm is smarter to which organization stands on a broader and deeper data foundation.

An organization that holds a wide range of data and also has a cyclical structure that returns that data through experiments seems highly likely to

gain an advantage in the next stage of new drug development competition. An organization that holds a reasoning foundation that crosses over diseases, genomes, and clinical data, rather than a calculator that hits a single target, will hold three practical tasks in one hand: generating new drug candidates, determining target priorities, and designing clinical trials. If Insilico Medicine advancing one candidate substance to phase 1 clinical trials in 30 months in 2022 was accepted as an exceptional achievement, the next task placed before organizations equipped with foundation models is to reproduce that level of speed simultaneously across multiple diseases.

Just as a chef who only cooked kimchi jjigae needed a long period of training to become a chef equipped with comprehensive basic skills, a time of training to gather data and verify it through experiments is still necessary before a life science foundation model leads to actual prescriptions beside a hospital bed.

In October 2025, the Ministry of Science and ICT selected a consortium led by Lunit, a medical artificial intelligence company, as the managing organization for developing a foundation model covering the medical science field. Here, twenty-three institutions participate, including companies like Kakao Healthcare and SK Biopharmaceuticals, research teams from KAIST and Seoul National University, and nine hospitals. They plan to complete the model by September 2026 by tying together multiple layers of medical science data, from molecules and proteins to clinical papers. The Ministry of Health and Welfare also set up the Korea Pharmaceutical and Bio-Pharma Manufacturers Association as the general managing organization to create a K-AI new drug development model that connects the preclinical stage to the clinical stage, together with thirty-one institutions.

This work, divided and led by Seoul National University Hospital, Samsung Medical Center, and the Korea Research Institute of Bioscience and Biotechnology, uses a federated learning method that gathers only the results learned within each hospital and combines them into a single model, without

moving patient data scattered in different places at each hospital to a single server. This is essentially a Korean-style experiment attempting to solve the old problem of data being trapped inside different servers by using a method that only gathers knowledge without moving the data.

There have also been cases where the predictions made by a foundation model were immediately overturned on the experimental bench. A research team led by Professor Markus Lill at the Department of Pharmaceutical Sciences, University of Basel in Switzerland, put AlphaFold 3 and several other structure prediction models to a mischievous test. They completely changed the amino acids in the binding site where the drug attaches or flipped the charge of the compound, making it actually impossible for the drug to bind to the protein, and then had the models solve the same problem again.

Looking at the results published in the international academic journal Nature Communications in October 2025, in over 40 percent of 285 complexes, the models produced the exact same answers as before, as if the binding sites had not changed. The research team explained regarding these results that the models did not understand why the drugs bind to the proteins, but rather just repeated the pictures they had seen in the training data exactly as they were.

The trend of training data and computational resources concentrating in a small number of companies seems to be a problem that cannot be fully solved by releasing weights alone. This is because the starting line itself is different between those who download and fine-tune weights and those who gather a massive amount of base sequences and clinical records in the first place and train models from scratch. Since companies that disclose both the composition and source of the original data used when training models are rare, it is also not easy for outside researchers to verify what the models have actually learned. Unless the public sector steps forward, like Korea's federated learning method where the government gathers budget and ties together the data of hospitals and universities, this concentration seems likely to widen

further in the future.

1.3

Multi-omics Integration and Discovery of Novel Biomarkers

The phase 2 TRIUMPH trial, conducted at various hospitals in Japan, targeted patients with metastatic colorectal cancer who had a significant amplification of the human epidermal growth factor receptor 2, or HER2 gene. These patients were first screened through a circulating tumor DNA screening study called GOZILA, which was a liquid biopsy method that detected the amplification of the target gene with a single bottle of blood instead of removing tissue again. This essentially opened the door to targeted therapy even for patients from whom it is difficult to remove tumor tissue again.

The researchers inputted the tissue slides of 143 patients who received the prescreening test and 30 patients who participated in the actual trial into an artificial intelligence pathology reading program. This program used software called SCOPE HER2 and SCOPE IO made by the Korean company Lunit, which simultaneously counted how darkly the cancer cells were stained with the HER2 dye and how densely immune cells and fibroblasts were clustered around the tumor in a single slide photograph. When compared to the judgments of pathologists, the accuracy of this program was 86.7 percent, and it was 100 percent accurate in picking out the level 3 patients with the darkest staining intensity.

The patient group in which more than half of the cancer cells were stained at that level 3 had a longer progression-free survival time of 4.4 months versus 1.4 months compared to the patient group that did not, and the overall survival time widened to 16.5 months versus 4.1 months. The results of this exploratory analysis were published in the international journal JCO Precision Oncology in 2025.

To properly understand these results, we must first explain the two words target and biomarker. A target refers to a protein or gene that causes a disease. You can think of it as a lock. Medicine is a key carved to perfectly fit that lock, and the door that caused the disease closes only when the key enters the keyhole and turns. A biomarker, on the other hand, is a signal that tells us what is happening inside the body right now.

Like a traffic light at a crosswalk or a warning light on a car dashboard, measuring the amount of a specific substance in a single drop of blood or a single tissue sample allows us to guess in advance what state the body is in, what medicine will work, and how the disease will progress in the future. The HER2 staining intensity and immune cell density read by the artificial intelligence in the TRIUMPH trial were exactly the biomarkers that acted as those traffic lights.

It is not uncommon for human pathologists to make different judgments on the same slide. The boundary between 2+ and 3+, which divides the HER2 staining intensity by eye, is vague, so these vague cases often had to undergo a fluorescence in situ hybridization test again. The artificial intelligence reading program transforms this boundary, which the human eye easily misses, into continuous numbers of staining intensity and cell ratio, greatly reducing the variation where results fluctuate depending on the judge. Of course, this does not mean that these scores replace the pathologist. It is more of an auxiliary tool that first filters out fine signals that are easy to miss before the pathologist makes a final judgment.

Cancer or degenerative brain diseases like dementia rarely occur because a single gene is broken. Hundreds and thousands of genes and proteins are intertwined and only reveal themselves outwardly as a disease in specific combinations. We cannot untangle this tangled skein with a single biopsy or a single genetic test. Therefore, researchers began to gather various technologies that look into the body at different layers. These are genomics,

transcriptomics, and proteomics. Genomics reads the entire DNA sequence inside a cell with next-generation sequencing equipment. You can think of it as a blueprint used when building a building. Once a blueprint is drawn, it rarely changes.

Transcriptomics corresponds to the work orders that the cell sends down day by day based on that blueprint. It measures how much of which gene is currently turned on using a technology called RNA sequencing. This is the reason why a liver cell and a nerve cell work differently even though they have the same blueprint. Proteomics goes one step further and individually checks the workers who actually carry bricks and drive nails at the site, namely the proteins themselves, with a mass spectrometer. Even the same protein moves completely differently depending on whether a small tag like a phosphate group is attached to it, and this precise proteomic analysis, which even examines these post-translational modifications, tells us much more closely the locations where medicine can actually step in along the signal transduction pathway.

Since medicines usually work by directly attaching to proteins rather than genes, knowing exactly how many of these workers there are and in what shape they are holding hands with each other is vital in new drug development. The problem is that it is very tricky for a person to lay these three data sets side by side and compare them because their measurement methods, units, and scales are all different.

When metabolomics and epigenomics are added to this, the picture becomes even denser. Metabolomics is the work of measuring the concentration of small molecules like glucose or amino acids to see what the cell actually eats and what it puts out, and epigenomics examines the changes that act as a switch to turn a specific gene on and off by attaching chemical tags to it while leaving the DNA sequence itself as it is.

Human genes are known to number just over 20,000, but the number of combinations these genes create as they go through the transcriptome, proteome, and metabolome stages is several orders of magnitude greater. It is nearly impossible from the start for a person to manually compare these vast combinations, and it is right here that the role of artificial intelligence, which swallows the multi-omics data whole to find patterns, grows larger.

This is where artificial intelligence steps in. Deep learning models like variational autoencoders or transformers take in genomic, transcriptomic, and proteomic data all at once and find hidden rules that are invisible to the human eye. Instead of forcing the three data sets to be averaged or laid side by side, this method creates a new coordinate space inside the model and moves each patient or cell to a point in that space. When we examine the distance and arrangement of the moved points, tangled relationships that could not be caught by traditional statistical techniques are revealed, for example, a relationship where a single gene and a single protein cause problems together only under specific conditions.

Market research firms project that the multi-omics analysis market has grown from \$2.81 billion in 2025 to around \$3.2 billion in 2026, and will grow by over 15 percent annually through 2035.

The work of identifying targets became more refined with the emergence of single-cell RNA sequencing (scRNA-seq). Previously, an entire tissue sample was ground up and the gene expression of all cells within it was measured in aggregate. This was like drawing an average face from a photograph of many people mixed together; in reality, the minority cancer stem cells or exhausted immune cells hidden within would disappear, buried in the statistics.

Single-cell analysis, by contrast, labels each individual cell and measures gene expression as if conducting separate interviews. In places like tumor tissue where normal cells, immune cells, and mutated cells are mixed together, there is no way to understand the different situations of each cell type without this

approach.

Researchers read tens of thousands of cells one by one in this manner, tracing the temporal flow from pre-cancerous cells through finely subdivided subpopulations of cancer cells to exhausted immune cells that have lost their function. Between 2025 and 2026, Illumina is accumulating single-cell transcriptome data reaching 20 petabytes in a single year through its Billion Cell Atlas initiative, and ultimately aims to build a database of 5 billion cells.

Single-cell foundation models trained on this accumulated data can name specific cell types and even predict how cells will respond to particular stimuli.

Artificial intelligence trained on single-cell data has also begun joining forces with literature-based language models. A framework called ELISA released in 2026 combines models like scGPT that understand single-cell gene expression with literature search functionality based on BioBERT, so that when researchers find a gene that is particularly active in a specific cell type, they immediately see related paper references. As experimental data and literature knowledge interlock this way on a single screen, it has become possible to visually confirm new biomarker candidates and immediately verify their validity.

But knowing the identity of each individual cell does not end the story. Cell function varies depending on where the cell is located. Spatial transcriptomics maps which genes are active at each location without grinding up the tissue, leaving it in its original form. It is like having a map that shows not just a list of residents' names, but which apartment in which building each person lives. Technologies like 10x Genomics' Visium or multiplexed error-robust fluorescence in situ hybridization (MERFISH), which applies multiple fluorescent labels simultaneously, draw this map with single-cell precision.

This field underwent rapid transformation between 2025 and 2026. In August 2025, 10x Genomics released Xenium Protein, opening the way to measure RNA

and protein simultaneously in the same cell and the same tissue section. In April 2026, it announced a new device called Atera, which covers an area four times wider than the existing Xenium using twice as many slides, measures the entire genome down to single-cell resolution, and uses graphics processing power to reduce analysis time from hours to minutes.

From data obtained with that equipment, location-specific biomarker candidates were newly identified, such as ODAM concentrated in hypoxic tumor tissue and C21orf58 concentrated in highly invasive dividing cells.

Artificial intelligence overlays single-cell data and spatial data together to calculate the actual physical distances between tumor cells, surrounding stromal cells, and immune cells, as well as the signals they exchange. In a study published in 2025, a graph neural network trained on spatial transcriptome and spatial proteome data achieved over ROC-AUC 0.90, an indicator of prediction accuracy, for predicting response to immune checkpoint inhibitors, and revealed that the region around the tumor enclosed by thick extracellular matrix was particularly prominent in patients who did not respond to the drug.

The immune cell density that Lunit's program read in the TRIUMPH trial is no different from this approach. In December 2025, Lunit published a paper in an international journal from a study of 291 cholangiocarcinoma patients and 309 slides showing that HER2 interpretation results matched pathologist consensus at 83.5 percent, and at the 2026 American Society of Clinical Oncology conference presented five studies, and at the American Association for Cancer Research conference presented six new studies including non-small cell lung cancer research conducted with Ajou University and Agilent.

Macrophages in the tumor microenvironment are also a focus of artificial intelligence. Macrophages are cells whose nature is not fixed, and can switch direction to either support tumors or fight against them. One spatial transcriptomics study revealed that two signaling molecules, CXCL16 and

SPP1, regulate this directional switching in macrophages, and reported that two drugs already used for other purposes, sirolimus and trichostatin A, emerged as candidates for targeting this signaling pathway.

This drug repurposing approach, which reuses drugs whose safety has already been confirmed for other diseases instead of designing new compounds from scratch, greatly reduces development time.

All the data examined so far are numbers directly extracted by laboratory equipment from cells and tissues. However, much of the knowledge about diseases that humans suffer from is not contained in numbers but in sentences within papers, patents, and clinical trial reports. It is simply impossible for people to read and organize the millions of papers poured out each year. This is why language models like BioBERT and PubMedBERT, pre-trained on vast amounts of biomedical literature, have emerged.

These models read a single sentence and automatically pick out which gene, disease, or drug names appear in it, and even extract how they relate to each other. A model called reguloGPT goes one step further and turns regulatory relationships, such as whether a gene suppresses or boosts the expression of another gene in a single sentence, into an organized table.

Gathering the relationships extracted from literature and the genomic, transcriptomic, and proteomic data obtained in the laboratory together creates a massive knowledge graph. It is a tightly woven map like a spider web, where genes, proteins, diseases, and drugs each become a dot, and the relationships between them become the lines connecting the dots.

Researchers apply techniques like graph neural networks or knowledge graph embeddings on this map to find which genes attract an unusually large number of lines within the interwoven network. Even relationships between drugs and targets that have not yet been reported in papers can be anticipated in advance on this graph.

Researchers call this kind of graph analysis disease network analysis. Just as you can be introduced to someone through a friend's friend without making friends directly, a graph neural network finds hidden links between gene A and gene C simply because both are frequently tangled with gene B, even if they have never appeared side by side in the same paper. KGDRP, a knowledge-guided graph learning technique published in 2025, boosted prediction performance by 12 percent in both drug response prediction and target prediction tasks, even for new diseases with almost no previously known data.

This means that artificial intelligence can present target candidates from the beginning even in rare diseases where literature or clinical data is still poor, and it appears it will significantly reduce trial and error in the early stages of new drug development.

Insilico Medicine's PandaOmics is a platform that commercialized this method early on. It weaves together gene expression, protein data, and literature-based knowledge graphs to prioritize new targets in several diseases, including hepatocellular carcinoma and idiopathic pulmonary fibrosis, and has verified them through actual experiments. In 2025, it refined a scoring system that jointly weighs reliability, commercial potential, druggability, and the clarity of the mechanism of action, releasing a new framework called TargetBench 1.0 and TargetPro, which covers 38 diseases across oncology, metabolic diseases, immune diseases, fibrotic diseases, and neurological diseases.

In the same year, Insilico Medicine joined hands with a world-renowned cancer center to find targets for gastric and esophageal cancer, and in 2026, it signed a second collaboration agreement worth 165 million dollars with China Medical System Group over brain disease targets selected by PandaOmics.

The case where this platform showed a clear outcome is rentosertib (development code name ISM001-055). PandaOmics found a protein called

TNIK while skimming through literature and omics data on fibrosis and aging, and this protein was not a candidate that had been eyed as a target for idiopathic pulmonary fibrosis treatments before then. In the phase 2a GENESIS-IPF clinical trial, which gathered 71 patients from 22 hospitals in China and compared a placebo and three dosages over 12 weeks, the patient group that received 60 milligrams per day saw their forced vital capacity (FVC) increase by an average of 98.4 milliliters, whereas the placebo group saw a decrease of 20.3 milliliters.

These results, published in Nature Medicine on June 3, 2025, were recognized as the first case in the industry to prove that a target and a drug found with artificial intelligence involved from start to finish actually produce effects in the human body, and this later led to phase 3 trials.

The methods to find targets also include synthetic lethality analysis. It is easy to understand if you imagine an airplane with two engines attached. Even if one engine stops, the flight can continue with the remaining one, but if both stop, it crashes. There are several such paired safety devices inside a cell, so even if one gene is ruined, another gene takes over the work in its place, and the cell survives.

However, cancer cells often grow with one gene already ruined, so if the remaining gene that forms the pair is blocked by a drug, normal cells remain fine, but only the cancer cells die. Artificial intelligence learns from pan-cancer multi-omics data and simulations that turn off genes one by one to find these pairs on a large scale.

Artificial intelligence models that find these pairs mostly use large-scale public databases called gene dependency maps as learning materials. The Cancer Dependency Map built by the Broad Institute in the United States is a prime example, and it is data that records through experiments which cell line dies when a certain gene is removed, turning off genes one by one across hundreds of types of cancer cell lines. When the genomic, transcriptomic, and proteomic

information of each cell line is overlaid on this, it is possible to statistically pick out pairs that only die exceptionally well in cell lines with specific gene mutations.

The fact that blocking an enzyme called PARP in cancer cells with a ruined BRCA gene selectively kills only the cancer cells first proved the possibility of this approach. Although there have not been many pairs that continued into clinical trials afterward, new candidates emerged one after another between 2025 and 2026. The relationship where blocking the PRMT5 enzyme kills the cancer cell in cancer cells missing the MTAP gene was confirmed, so IDEAYA Biosciences submitted a clinical trial plan for a PRMT5 inhibitor called IDE892 in September 2025, and started a phase 1 trial for patients with MTAP-deleted lung cancer in the fourth quarter of the same year.

Bristol Myers Squibb's BMS-986504 led to a Phase 2 trial recruiting 60 patients with MTAP-deficient pancreatic cancer. In pancreatic ductal adenocarcinoma research, a pipeline combining single-cell spatial transcriptomics and proteomics, competitive endogenous RNA network analysis, and transformer-based deep learning called SELFormer identified a death receptor called TRAIL-R1, produced by a gene named TNFRSF10A, as a new target. Computer simulations confirmed that temsirolimus, already approved as a treatment for renal cancer, binds to this receptor.

The practical use of discovered biomarkers in clinical settings can be seen in KEYNOTE-002 and KEYNOTE-006, large-scale clinical trials of pembrolizumab, a melanoma treatment. Researchers extracted twenty-five features showing tumor size and shape changes from CT images of 575 patients, then used a random forest algorithm to select four imaging biomarkers: two related to size and two related to shape changes.

This combination of four biomarkers predicted patient overall survival quite accurately at the six-month point after treatment began. The fact that biomarkers could be extracted using only CT images that had already been

taken, without a single drop of blood or a tissue sample, demonstrates how easily this technology can integrate into hospital practice.

Target candidates identified by disease networks or multi-omics models on a computer screen do not become drugs by themselves. They must go through a stage where preclinical experiments using cell lines, organoids, or animal models confirm whether suppressing that target actually improves the disease. Only candidates that pass this stage move on to clinical trial design. Exploratory analysis approaches, like re-examining tissue slides and imaging data from large-scale clinical trials already underway for other purposes - such as the TRIUMPH trial or KEYNOTE trials - using artificial intelligence can rapidly validate biomarker candidates without designing new clinical trials from scratch, thereby saving considerable time and cost.

Only after validated biomarkers receive companion diagnostic approval can they actually be used in hospital clinics for real patients.

Confirmed biomarkers are now changing clinical trial design itself. In the past, patients were randomly divided, and whether one treatment was better than another was determined only after the trial ended. Now, adaptive clinical trial designs are becoming more common, where biomarker values are monitored in real-time during the trial, allowing patient groups with poor responses to switch to different treatments while those with good responses continue as they are.

Starting the main trial only after pre-selecting patients using circulating tumor DNA, as in the TRIUMPH trial, is one example of this trend. By reducing patient confusion and reaching conclusions more quickly with fewer participants, biomarker discovery appears to have the power to change the pace of drug development itself.

For discovered biomarkers to lead to actual clinical prescriptions, they must pass through a gateway called companion diagnostics (CDx), which are test

tools paired with drugs to ensure that specific drugs are used only in patients with specific biomarkers. The number of companion diagnostic tests approved by the U.S. Food and Drug Administration was 34 in 2017 and 44 in 2021, but as of March 2025, this had grown to 188 tests, with 45 related biomarkers. In May 2026, Natera's Signatera companion diagnostic for measuring recurrence risk using blood circulating tumor DNA in bladder cancer patients, and Guardant Health's Guardant360 Liquid companion diagnostic spanning breast cancer, colorectal cancer, and non-small cell lung cancer, received approval in succession.

The U.S. Food and Drug Administration gathered input through January 2026 on ways to transition established cancer companion diagnostic methods - such as next-generation sequencing-based tests or polymerase chain reaction-based tests - to a lighter approval process instead of the current stringent premarket authorization. However, cases where new drugs and companion diagnostics receive approval at the same time remain rare, so situations often occur in actual clinical practice where the drug comes out first and the test to designate it follows later, or vice versa.

Korean companies are increasingly following this trend. Arontier has established a platform connecting the entire drug discovery process, from REMEDY, which discovers new targets and candidate substances using multi-omics data, to DRUG DESIGNER, which designs small-molecule compounds, to PATH, which analyzes tumor microenvironments using pathology images. At the 2026 American Society of Clinical Oncology meeting, in addition to Lunit, domestic companies such as STcube, Tium Bio, ImmuneOnCia, and OnconiC Therapeutics are scheduled to present clinical data on artificial intelligence-based biomarkers and synthetic lethal-based anticancer drugs.

Despite all these achievements, there remain significant hurdles for artificial intelligence-based target discovery to take root as a routine tool in hospitals.

The most immediate problem is the bias in the data itself used for learning. The majority of genomic and clinical databases currently accumulated are skewed toward Western or specific developed-country populations, so when the same model is applied directly to patients of other races and regions, predictions often go awry. When data is skewed in one direction, the answers produced by models are bound to be skewed in that direction as well. This carries the risk of actually widening health disparities.

Another obstacle is that it is difficult for people to trace exactly why deep learning models produce certain answers. Multi-layered neural networks and approaches that mix multiple models excel in predictive performance but are difficult to examine internally, which is why they are commonly called black boxes. This kind of opacity is hard to accept in the face of prescriptions where patients' lives are at stake and before regulatory agencies deciding whether to approve new drugs.

That is why researchers are making efforts to use explainable artificial intelligence techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) together, translating how a model's decisions originate from specific genetic mutations or metabolic pathways into language that people can understand.

The gap in laws and systems also remains. Protecting personal information during the process of gathering and training on patient data scattered across multiple hospitals and multiple countries is the most sensitive issue. To comply with the General Data Protection Regulation (GDPR) of the European Union and the Health Insurance Portability and Accountability Act (HIPAA) of the United States, federated learning technology is increasingly spreading, where the model is trained only inside each hospital without moving the original data outside, and only the resulting values are exchanged. However, different data standards and encryption costs across countries still remain a stumbling block.

Disputes are also breaking out in various places around the world over who should be recognized as the inventor of targets found or biomarkers designed by artificial intelligence, and whether those rights belong to the company that created the algorithm or the hospital that provided the data. In response to these problems, the United States Food and Drug Administration and the European Medicines Agency are continuing to revise regulatory documents that clarify model transparency, quality control, and the locus of responsibility for safety after market release.

The industry expects that between 2025 and 2026, there will be more than ten clinical results for drugs that started from targets found by artificial intelligence like this, and as those records accumulate, it seems that the reliability of the approach itself, which finds targets by combining omics and literature data, will also increase along with them.

Looking at the flow that has been revealed while passing through 2025 and 2026, it seems that the combination of multi-omics and artificial intelligence has moved past the laboratory demonstration stage of showing off computational ability, and into the clinical field where it divides the survival periods of actual patients. The HER2 patient group of the TRIUMPH trial, the pulmonary fibrosis patient group that took lentosertinib, and the new inhibitor aiming at MTAP-deficient cancer are all results that were obtained only after tying together heterogeneous omics data and weaving knowledge from literature into a knowledge graph.

It is highly likely that the work of standardizing and connecting data that comes from different hospitals and different countries, and translating the basis for judgment of black box models into words that humans can understand, will determine the success or failure of the next few years. The current flow, in which a signal starting from a single tissue slide or a single drop of blood from one patient connects with the knowledge graph of a researcher on the other side of the planet, appears to be weaving anew the

foundation of the old work of new drug development.

The South Korean government is also carrying out a project to accumulate multi-omics data on a national level. The National Integrated Bio Big Data Project led by the Korea Disease Control and Prevention Agency has set the goal of gathering the genomic and clinical information of 1 million citizens into one place. In the first phase, which runs from 2024 to 2028, it decided to gather 772,000 people, combining patients with rare diseases, patients with severe diseases, and general participants. Citizens who wish to participate go to one of the 38 recruiting institutions nationwide, fill out a consent form and a questionnaire, leave a specimen such as blood or urine, and then hand over their clinical information together.

The project team explains that the materials gathered this way will become the foundation for finding targets and biomarkers that fit South Koreans. The genomic data gathered this way is scheduled to be opened to researchers in turn starting in 2026, and as of the end of April 2026, the number of participants had grown to 165,722 people.

However, the data of patients with severe diseases, which is actually of great use in developing new drugs, falls far short of the target. As of the end of April 2026, only 11,274 patients with severe diseases had been gathered out of the target of 140,000 people, barely passing 8 percent. Only 4,476 patients with rare diseases were gathered out of the target of 47,000 people, stopping at 9.5 percent. The project team revealed the reason, stating that there is a shortage of research nurses to collect specimens and persuade patients, and the motivation for medical staff to step up and participate is also weak.

When thinking about the survey results showing that the proportion of European populations among the data used in genomic research worldwide exceeds 90 percent, the weight of this project, which will fill in the genetic information of South Koreans and other Asians, is never light. Nevertheless, if the materials of the people who are actually suffering from illnesses

accumulate this slowly, it seems that the opportunity to correct the data skew will also not be grasped for a long time.

For discovered biomarkers to lead all the way to hospital prescriptions, they must cross three thresholds in order. The first is analytical validity, which looks at whether the testing tool accurately and consistently catches the target it aims for. The second is clinical validity, which checks how well that target tells in advance about the patient's treatment response. The third is clinical utility, which measures whether the patient's treatment and results actually improve when this test is used. The United States Food and Drug Administration recommends inserting a companion diagnostic test together starting from the time when the phase 3 clinical trial of a new drug candidate is designed.

This is because doing so increases the possibility that the drug and the test will be approved side by side on the same day. If the test preparation is delayed, a situation happens where the drug comes out first, but prescriptions are delayed because there is no tool to single out the patients who should use that drug. The South Korean Ministry of Food and Drug Safety is also keeping a similar principle by requiring separate plan approval to conduct clinical performance tests with companion diagnostic medical devices.

PART 2

AI-Driven Drug Design and Molecule Discovery

2.1

3D Protein Structure Prediction and Structure-Based Drug Design (SBDD)

On November 30, 2020, hundreds of structural biologists from around the world gathered in front of their computer screens. It was the moment when the scoring results of AlphaFold 2, an artificial intelligence created by Google DeepMind, were released at the 14th International Protein Structure Prediction Competition, commonly known as CASP. The score displayed on the screen exceeded 90 points. A single computer had predicted the three-dimensional shape of a protein, something that researchers could barely confirm after months or, at most, years of work using X-ray crystallography or cryo-electron microscopy, with almost the same accuracy as experimental results.

That day, a problem that had tormented biologists for nearly half a century was essentially solved. In the early 1970s, the principle that amino acid sequences determine the three-dimensional structure of proteins had already been discovered, but the task of calculating the actual shape based solely on the sequence remained unsolved for nearly half a century afterward.

Proteins are substances made of small components called amino acids linked together in a long chain. When a chain is first created inside a cell, it hangs loosely like a ball of thread, but in water it quickly and precisely folds into a unique three-dimensional shape. It is easy to understand if you think of origami. Depending on the folding order and angle, the same square piece of paper can become a crane or a boat.

Similarly, proteins fold in a specific way depending on the sequence of amino acids, and if the shape is even slightly wrong, the protein cannot perform its original function or attaches to the wrong place and causes problems. This is why we know of several diseases, such as Alzheimer's disease and Parkinson's

disease, caused by proteins that misfold and aggregate.

The first piece of information pharmaceutical companies need to know when designing a new drug is this three-dimensional shape. Only by knowing exactly what shape the disease-causing protein has can they design small molecules that bind to the protein and block or enhance its function. This approach of designing drugs based on the three-dimensional structure of a target protein is called structure-based drug design, or SBDD for short. In the past, to obtain this structure, researchers had to use X-ray crystallography, in which proteins are crystallized and X-rays are used to read diffraction patterns, or cryo-electron microscopy, commonly called cryo-EM, in which samples are frozen at extremely low temperatures and electron beams are applied while images are superimposed.

Both methods required skilled personnel, expensive equipment, and time ranging from months to years. It was not uncommon for the process of crystallizing proteins to not proceed as planned, requiring a complete change in approach.

The predictive power that AlphaFold 2 demonstrated on the CASP stage shook this old way of working at its roots. In May 2024, Google DeepMind and its subsidiary Isomorphic Labs took a step further by releasing AlphaFold 3, the sequel. AlphaFold 3 does not simply predict the shape of a single protein; instead, it predicts with a single neural network the three-dimensional structure of an entire complex containing DNA and RNA carrying genetic information, small molecular ligands that are drug candidates, metal ions, and even post-translational modifications.

By combining an evolved Evoformer module with a diffusion model, it starts from a state where atoms are scattered like clouds and gradually removes noise to find an arrangement close to the correct answer. When it was released, the academic community evaluated that the accuracy of predicting interactions improved by more than 50 percent compared to existing methods.

Although the term diffusion model is unfamiliar, the principle is easy to understand if you think of the crackling screen of an old television. Just as a clear picture gradually emerges as the channel tunes in on a screen that initially shows only static, AlphaFold 3 starts from a noisy state where atoms are randomly scattered and gradually removes noise step by step to complete a picture close to the actual molecular arrangement. As this process is repeated tens of times, the initially hazy cloud of atoms narrows into a clear three-dimensional structure where proteins and ligands interlock.

These predicted structures have accumulated to become a public asset known as the AlphaFold Protein Structure Database. As of 2025, with its newly redesigned interface, the database contains over 214 million protein structures, making it possible to search for proteins from virtually all species of life, including humans, with just one search. In fact, Chinese researchers used this database and AlphaFold to identify 315 abnormal structures appearing in patients with uveal melanoma and gained clues to new therapeutic targets. This scale would have been unimaginable in the days when it took years to determine a single protein structure in a laboratory.

The tangible change is first seen in speed. Before AlphaFold 2 appeared, confirming the target protein structure, which is the first gate in discovering new drug candidates, was a task that took months. After AlphaFold 2 came out, this task was reduced to a matter of days, and the time saved is being reinvested directly in the next steps of docking and simulation.

When AlphaFold 3 was first released, there was significant disappointment because researchers could not directly download and run the code; they could only input sequences into the AlphaFold server operated by Google and receive results. Starting in 2025, Google DeepMind expanded access by allowing the inference code to be downloaded and model weights to be requested for non-commercial academic research, and doubled the daily processing capacity of the AlphaFold server. Meanwhile, efforts to create

alternatives were also active.

Chinese company ByteDance released ProteinNix, an open-source model that completely replicated AlphaFold 3, and Columbia University's AlQuraishi Lab published OpenFold3-preview under the Apache 2.0 license, which permits both academic and commercial use for free. What was once a technology monopolized by a single company has spread over several years into an open-source ecosystem shared by multiple research teams.

If AlphaFold 3 competes on precision, ESMFold developed by Meta competes on speed. Most structure prediction artificial intelligences calculate structures only after going through a multiple sequence alignment process, which involves finding similar proteins across multiple species and lining them up side by side. This process alone takes a considerable amount of time just to search the databases. On the other hand, ESMFold is based on the large language model ESM2, which has pre-learned hundreds of millions of protein sequences, and outputs a structure immediately when just one sequence is inputted, without multiple sequence alignment.

The strengths of ESMFold stand out in cases where there are no suitable targets for comparison, such as orphan proteins with almost no evolutionarily similar relatives or newly discovered viral proteins, and it is widely used in the field for large-scale screening and new enzyme discovery. Based on this technology, Meta has also predicted and released on a large scale the structures of proteins that exist in nature but are not yet known.

Another pillar supporting the protein structure prediction ecosystem alongside AlphaFold is RoseTTAFold, created by the research team of Professor David Baker at the University of Washington. Among them, the recent version, RoseTTAFold All-Atom, predicts the structures of complexes where proteins, nucleic acids, and small molecule ligands are mixed together, similar to AlphaFold 3. The true specialty of Professor Baker's research team lies not just in guessing structures, but in designing entirely new proteins that bind to

desired molecules from a blank slate. Tools in the RFDiffusion family, which apply diffusion models, take on that role.

RoseTTAFold was first announced in 2021 with a triple neural network structure that simultaneously processes amino acid sequence information, distance information between atoms, and three-dimensional coordinate information in different neural network pathways, and then has them exchange information with each other. Released around the same time as AlphaFold 2, this model was evaluated as falling slightly short of AlphaFold in prediction accuracy, but it possessed the practical strength that it required relatively fewer computer resources for calculation, allowing individual laboratories to run it.

This tool took another leap forward on December 3, 2025, under the name RFDiffusion3. This model, which handles up to the atomic level, can design not only proteins but also proteins that interlock with DNA or various ligands, and its inference speed has become nearly ten times faster than the previous version. The announcement that it could design new enzymes inside a computer with performance approaching enzymes refined over a long time in nature caused a considerable stir in academia. This is because it means the cycle of developing new industrial enzymes, which used to take several years, can be changed to a method where candidates are first drawn inside the computer and only go through verification in the laboratory.

Before that, the same laboratory published a paper in the international journal *Nature* succeeding in designing binders that attach to intrinsically disordered proteins, which are difficult to handle because they have no fixed shape, and also released the results of the RFantibody study, which designs the entire heavy and light chains of antibodies solely with a computer. Because antibodies are already being used as primary drugs for the treatment of various intractable diseases, the fact that the framework of an antibody can be designed starting from a computer screen has the potential to change the

biopharmaceutical development strategy itself. This means that the center of gravity is shifting from artificial intelligences that read structures to artificial intelligences that create structures as desired.

The area that has long waited for such technology is precisely the domain called undruggable proteins. Proteins with no distinctively carved binding pockets, leaving no suitable handles for drugs to grab; intrinsically disordered proteins that constantly change shape inside cells; and binding sites between proteins that are so wide and flat that a single small molecule cannot possibly block them - these targets have long been excluded from lists of new drug development candidates. This is because it was difficult to make such proteins the targets of design with just a single static crystal structure photograph. Generative artificial intelligence, which has learned millions of sequence and structure data, is overcoming this wall in a different way.

Although the name intrinsically disordered protein may be unfamiliar, the meaning is not difficult. Unlike most proteins that fold neatly into a fixed shape like origami, these proteins continue to sway in water without a fixed form. Because they have no shape to hold onto, they effectively had no suitable keyholes to try fitting into, and thus they remained outside the interest of new drug developers for a long time. In 2026, a benchmark called ShallowBench also appeared, which gathers only targets with shallow pockets to compete on performance, beginning to closely distinguish the abilities of artificial intelligence models that handle undruggable proteins.

Proteins have a back door in addition to a front door. If the active site that drugs commonly target is the front door, the allosteric site corresponds to the back door on the protein surface away from the active site. When a small molecule gently touches this back door, the three-dimensional shape of the entire protein changes, regulating the activity of the enzyme or signal transduction. Tools like DeepAllo, which jointly analyze features extracted by protein language models and the appearance of binding pockets, find the

allosteric sites of kinases or G protein-coupled receptors that are not revealed by a single crystallography photograph. This means that even proteins that could not be touched because their front doors were firmly closed can have their functions selectively regulated by using their back doors.

Actual results have also followed. The structure of the epidermal growth factor receptor, or EGFR mutation, and the information on the three-dimensional shape changes of the KRAS protein revealed by AlphaFold were actually used to refine the efficacy of sotorasib and erlotinib, precision targeted therapies used for lung cancer and breast cancer patients. For a long time, KRAS was a target notorious as a protein with a smooth surface and no handles. The fact itself that KRAS is a gene related to cancer was discovered a long time ago, but because no tangible binding pocket could be found, it could not shake off the tag of being an undruggable gene for decades. As structure information becomes denser, methods to slowly tackle even these proteins are being found.

Recently, research has also been active in going one step further from AlphaFold-based models by predicting multiple forms that proteins can shift between all at once, rather than a single correct structure. Taking into account that proteins take on different shapes before and after binding, laying out multiple snapshots rather than a single photograph is an attempt that approaches actual movement more closely.

However, the structure that AlphaFold produces is much like a single photograph capturing a single moment. Real proteins are not the frozen forms seen in photographs but living structures constantly trembling and twisting while surrounded by body heat and water molecules. It is the same principle as needing to watch video rather than still photographs to see true motion. The technique that reproduces this motion through calculation is molecular dynamics, or MD simulation for short.

MD continuously updates the position and velocity of each individual atom at very short time intervals according to Newton's laws of motion, tracking how proteins shake over time scales ranging from picoseconds to milliseconds, and how surrounding shapes change in response at the moment a ligand binds.

If AI-based molecular docking serves the role of quickly taking multiple snapshots of binding, MD simulation takes on the role of verifying how long and how stably those snapshots are maintained in the actual in-vivo environment. MD simulation captures not only the induced fit phenomenon where protein shape changes along with ligand binding, but also the subtle effects created as water molecules enter and leave the binding site. It is not uncommon for candidate compounds that looked promising based only on docking results to have their binding completely dissolved within hours during the MD simulation stage. Only after passing through such verification does a candidate compound earn the qualification to move to the next stage.

The movement of electrons at the moment when chemical bonds are newly formed or broken - that is, the moment when an enzyme catalyzes a reaction or a drug attaches to a target protein with a covalent bond - cannot be calculated precisely even with MD simulation. To deal with such a fine-scale world of electrons, quantum mechanics equations are needed. However, calculating an entire protein using quantum mechanics would take time that is difficult even for supercomputers to handle. Therefore, a hybrid calculation technique of quantum mechanics and molecular mechanics, called QM/MM for short, is used: calculating only the region around the active site where binding actually occurs using precise quantum mechanics, while treating the remaining massive protein framework and water molecules using classical mechanics, which is computationally less expensive.

It is a similar concept to constructing a building not with uniform precision throughout, but using precision components only for door handles that hands directly touch, while building the remaining walls and columns with standard

materials.

Recently, computational methods combining quantum machine learning and graph neural networks have been significantly reducing the cost of QM/MM. Researchers at Pohang University of Science and Technology have developed a machine learning potential combining charge equilibrium principles with graph neural networks, presenting a method that approaches quantum mechanics-level accuracy while being much faster computationally. At a drug development conference in 2026, Schrödinger announced that it had unveiled a workflow connecting physics-based simulation and generative artificial intelligence, generating over 12,000 compound candidates in just 4 hours and 30 minutes, which drew attention.

This means that the reliability of in silico validation, which filters out compound toxicity and pharmacokinetic properties in advance within computers at the preclinical stage, is rising correspondingly.

As these various technologies accumulate, the procedure by which artificial intelligence identifies new drug candidates generally goes through four stages. First, researchers input the amino acid sequence of the protein that causes the disease. Next, artificial intelligence like AlphaFold predicts the three-dimensional structure of the protein from that sequence alone. Once the structure is in hand, molecular docking programs fit each candidate compound to that structure's binding site and select candidates that match well.

Finally, QM/MM calculations and MD simulation precisely recheck the electron movement around the binding site and stability over time, refining the final candidates. This journey, from starting with a single sequence to reaching validated candidate compounds, which took years in the past, is now being compressed into weeks or months.

In this workflow, the third stage, molecular docking, resembles trying keys one by one in a keyhole. If we think of the grooves carved into the surface of a target protein as keyholes, candidate compounds correspond to the task of finding the key that fits those holes. However, what differs in the drug development field is that a single keyhole must be matched simultaneously with keys chosen from millions or hundreds of millions of candidates. In the past, classical programs relying on physical energy functions and empirical scoring like AutoDock Vina or Glide handled this task.

Recently, artificial intelligence docking models such as DiffDock, SurfDock, Chai-1, and Boltz-1, combining deep learning, graph neural networks, and diffusion models, have appeared in succession. After learning the surface shape and electrostatic properties of target proteins, they find flexible binding poses much more quickly. Virtual screening time to sift through libraries of hundreds of millions of compounds has been greatly reduced as a result.

The situation is somewhat different in protein-protein docking. Traditional programs based on physical laws like ZDOCK finely reflect the three-dimensional structural changes occurring during the binding process, whereas several deep learning models merely perform static shape matching assuming the forms before and after binding are fixed, sometimes distorting the actual binding environment. However, the recent atmosphere differs in docking connecting proteins with small molecules.

A new benchmark called PoseX, released in 2025, tested unfamiliar combinations of proteins and ligands not present in the training data. Here, some artificial intelligence models showed results surpassing traditional physics-based methods. It is still premature to say that either approach has achieved complete superiority.

In June 2025, Boltz-2, which was released together by the Massachusetts Institute of Technology and the drug development company Recursion, added another layer to this trend. This is because it is the first co-folding model that

not only guesses the structure but also predicts the strength of the bond, that is, the binding affinity, within the same model. It gathered interest from both academia and industry at the same time by receiving evaluations that it produces an accuracy matching sophisticated physics-based calculations called free energy perturbation at speeds up to a thousand times faster.

Boltz-2 was initially released as open source, but the follow-up version Boltz 2.1, which came out in the first half of 2026, transitioned to a closed type provided only in the form of an API. It is an example that shows how the flow of technology is shifting from openness back to commercialization.

The tug-of-war surrounding the scope of disclosure seems to be a trend that runs through this entire field. While AlphaFold 3 opened its doors wider to academia and alternatives like Protenix and OpenFold3-preview followed, Boltz-2, which actually predicted even binding affinity, locked its doors in its follow-up version. Just as much as creating a single good prediction model, deciding who to open the model to and how much seems to be becoming a factor that determines victory or defeat in this ecosystem.

Behind the splendid achievements, unresolved flaws are also clear. When examining the binding pose prediction results of AlphaFold 3 with the strict verification standard called PoseBusters, it is just over 76 percent on the basic dataset, and even if the grading standard is loosely changed, it stays around 84 percent. The really serious problem lies elsewhere. Among the binding poses produced by deep learning models, there are not uncommonly mixed structures that cannot exist in reality, such as atoms physically overlapping each other, or failing to distinguish optical isomers whose directions are opposite like a shape reflected in a mirror.

When facing a bond type similar to the training data, artificial intelligence shows surprising accuracy, but in multiple benchmarks, it has been repeatedly confirmed that its predictive power drops sharply in front of unfamiliar proteins or completely new compounds it has never seen before. Therefore,

the procedure of refining the binding pose produced by artificial intelligence once more with physics-based calculations to correct awkward structures is solidifying as a practical solution to raise accuracy.

A place that aimed squarely at this problem is Isomorphic Labs, which launched by spinning off from Google DeepMind in 2021. The company announced that IsoDDE, which it released in February 2026, showed an accuracy about twice as high as AlphaFold 3 in truly unfamiliar ligand binding cases where the sequence similarity with the training data did not even reach 20 percent. Isomorphic Labs has joined hands with large pharmaceutical companies like Novartis, Eli Lilly, and Johnson & Johnson to develop new drug candidates together, while also cultivating its own candidate substances in the fields of oncology and immunology.

At the Davos Forum in January 2026, Demis Hassabis stated that the schedule for the first clinical trial, which was initially set for 2025, was delayed to the end of 2026, but he also made it clear that patient administration is not far off. Early in 2026, it was also announced that Johnson & Johnson engaged in an in-depth collaboration to design new inhibitors that block the bonds between proteins based on AlphaFold 3.

Related research is also following in South Korea. Researchers at KAIST introduced an artificial intelligence that goes beyond whether a drug attaches to a target protein and also predicts whether it actually operates as desired after attaching. It is an attempt that raised the standard of existing models that only looked at whether it binds or not by one step. The startup Galux, founded in 2020 by Professor Chaok Seok of the Department of Chemistry at Seoul National University, designs proteins with artificial intelligence that uses the laws of physics as its framework, based on a history of achieving top-tier results in international protein structure prediction competitions even before AlphaFold appeared. It conducted a collaboration to develop anticancer proteins together with LG Chem, and the investment funds it has gathered so

far reach 68 billion won.

Simplex, which opened its doors in South Korea around the same time, is also worth paying attention to. This company, founded by CEO Sungjin Cho, who has developed new drug research platforms for over 20 years at global pharmaceutical companies GSK, BMS, and Amgen, puts its strength in filtering out candidates that have a high possibility of failing even before they enter the clinical stage in advance, featuring an explainable artificial intelligence that also shows the reasons why the prediction results came out.

Experimental techniques that directly confirm structures have also joined hands with artificial intelligence. Artificial intelligence that automates the tasks of picking out, classifying, and reassembling particles in three dimensions has been introduced to RELION or cryoSPARC, which are software programs that process cryogenic electron microscope images, reducing the processing time that used to take weeks to a few hours. In front of density maps that are blurry in shape and difficult to interpret, the prediction structures produced by AlphaFold serve as reference maps and help the process of putting the models together. It has changed into a structure where the blurry raw data created by experimental equipment and the map close to the correct answer drawn in advance by artificial intelligence match each other back and forth.

The fact that large pharmaceutical companies including Pfizer, Novartis, and AstraZeneca have brought AlphaFold and RoseTTAFold into their in-house new drug development pipelines is also in touch with this trend. The sight of the structural biology department sharing an office with the computational team running prediction models is no longer unfamiliar either. A circular structure where experiments and calculations take turns verifying each other continues to reduce the time it takes to figure out a single target protein.

At the foundation of all these predictions, the opacity unique to deep learning still remains. It is not yet easy for a researcher to follow line by line and explain

the inner logic of that judgment, or why the neural network picked a specific atom arrangement as the correct answer. Because of these characteristics where it is difficult to retrace on what physical basis a single binding affinity score came out, the calculation results are being treated not as final answers but as candidate lists that need to be narrowed down through experiments.

The event that set an official milestone for this entire flow was the 2024 Nobel Prize in Chemistry. The Royal Swedish Academy of Sciences awarded half of that year's Nobel Prize in Chemistry to David Baker, and the other half jointly to Demis Hassabis and John Jumper. Baker was recognized for designing novel proteins not found in nature, while Hassabis and Jumper were recognized for predicting protein structures using only amino acid sequences. The very fact that the prize money of 11 million Swedish kronor went to an AI company CEO and researchers, and to a university professor, rather than a single laboratory researcher, was unprecedented.

The scientific community recognized the achievement of solving long-standing problems in chemistry and biology through computer calculation as the highest authority, and afterward the entire pharmaceutical industry accelerated in accepting AI-based structure prediction not as a peripheral experiment but as a standard procedure in drug development. A methodology that had for some time been viewed only as an experimental attempt in computer science had risen in just a few years to stand among the ranks of established science receiving the Nobel Prize, and this recognition is also drawing more young researchers into this field.

As the method of obtaining structures expanded from a single experiment to multiple layers of calculation, the bottleneck in the early stages of drug development is no longer the lack of structure. It appears that the bottleneck has shifted to determining which of the many predicted structures and binding candidates to place on the experimental bench first.

Inputting a single protein sequence and receiving predicted structures in a short time, testing hundreds of millions of compounds against that structure, and validating promising candidates through physical calculation have now become routine tasks for multiple pharmaceutical companies and startups. However, whether the binding conformation completed on a computer screen actually remains the same inside the human body must ultimately be verified once more using traditional experimental methods such as surface plasmon resonance or X-ray crystallography. The faster this cycle of prediction and experiment proceeds, the gradually shorter becomes the list of proteins that had long been considered untouchable.

Getting the structure exactly right and measuring the strength of binding on that structure accurately are different problems. Even if AlphaFold or models of the Boltz series match the position of each individual atom almost exactly with experimental results, predicting binding affinity - which indicates how strongly and how long the binding is maintained inside the body - still remains difficult to answer. This is because binding affinity is a value that cannot be captured in a single still photograph. The degree to which water molecules enter and exit the binding site, the loss of freedom as the chain folds, and what charged state each amino acid and ligand assumes - all of these must be added and subtracted to arrive at a single value.

Boltz-2 showed performance matching physics-based calculations in some validations, but in other validations against proteins with continuously changing forms, it also showed performance similar to randomly selecting even compounds with actual activity. Considering the fact that even when measuring binding strength experimentally, an error of around 1 kcal/mol is introduced, no matter how sophisticated the calculation becomes, there is from the start a wall below which it is difficult to descend.

As these prediction tools became widely distributed, an entirely different concern also grew. When tools that design proteins not found in nature are

placed within anyone's reach, a path also opens for dangerous proteins to evade surveillance by changing only the sequence slightly. In May 2025, the US government issued an executive order instructing improvement of safety and security in biological research and mandating redrafting the screening criteria for synthetic nucleic acids used in government-supported research. Five months later in October, a research team led by Microsoft Chief Science Officer Eric Horvitz published the loopholes in that screening system to the international journal *Science*.

They substituted the sequences of 72 types of dangerous proteins including ricin and botulinum neurotoxin with publicly available AI design tools and created over 76,000 designs inside a computer, confirming the fact that a considerable number of them would have passed through the existing surveillance software that genetic synthesis companies had been using. Although it was an experiment inside a computer without actually synthesizing proteins, the research team worked with the International Gene Synthesis Consortium to improve the software, and even after that, about 3 percent of the modified sequences were still not filtered out.

The gap between the answers that calculation produces and the ability to verify whether those answers can actually be trusted appears to be a problem that repeats throughout this entire field. Whether in measuring binding strength or in filtering out dangerous sequences, the structure in which the ability to create runs ahead of the ability to verify has a strong likelihood of continuing for some time to come.

2.2

Deep Learning-Based Drug Candidate Generation (De Novo Design)

In 2022, a computer in a pharmaceutical company laboratory in Shanghai, China, designed a single substance to cure a disease called idiopathic pulmonary fibrosis. Idiopathic pulmonary fibrosis is an intractable disease where lung tissue slowly hardens without a known cause, making it difficult to breathe. It is such a severe disease that the average survival time after diagnosis is only about three to four years, making the patients' wait for a new treatment desperate. The company that designed this substance is Insilico Medicine, the target was TNF, an enzyme that carries inflammation and fibrosis signals inside the cells, and the resulting substance was named ISM001-055.

The time it took from setting the target to entering the phase 1 clinical trial involving humans was thirty months, less than three years. Cases where artificial intelligence took charge from target discovery to molecule design and reached clinical trials were among the very few in the world at the time. In 2025, the results of the phase 2a clinical trial were published in the international journal Nature Medicine and at the American Thoracic Society conference. This trial, conducted with 71 patients gathered from 21 hospitals in China, met its goal of safety, and forced vital capacity, an indicator that measures how much air the lungs breathe in and out, also improved along with the increase in dosage.

On July 7, 2026, an announcement was made that it would enter a phase 3 clinical trial. A new drug candidate that humans had been searching for through decades of trial and error was drawn from a blank slate by a computer, and the results are being confirmed one by one in patients' bodies.

This substance later even received the International Nonproprietary Name of rentosertib. An International Nonproprietary Name is the official drug name granted by the World Health Organization, which also means that it has begun to be treated as a formal medicine used together by the whole world, moving beyond a single company's laboratory code name.

For a single new drug candidate to pass from target discovery to animal testing, it takes over four to six years and enormous amounts of money using traditional methods. According to estimates frequently cited in the pharmaceutical industry, the cost to bring a single new drug to the market exceeds 2 billion dollars. During this process, most candidate substances drop out in the middle, and people in the industry call this section the valley of death. Pointing to this trend, the industry also uses the term Eroom's Law. It is a name that reverses Moore's Law, which states that semiconductor performance steadily improves, meaning that the cost and time required to develop a single new drug actually continue to increase even as technology advances.

It is a long-accepted rule of thumb in the industry that the ratio of even those candidate substances that have reached the clinical trial stage surviving to final approval is less than one in ten. When artificial intelligence first entered this industry, the role it took on was quickly searching through warehouses stacked with billions of chemical compounds, like ZINC or Enamine REAL. The computer could scan through trillions of compounds in the blink of an eye to pick out promising candidates, but the substances found this way were, after all, nothing more than things someone had already made and placed in the warehouse. If the correct answer is not inside the warehouse, no matter how quickly you search, you cannot find the answer.

What deep learning has newly brought to new drug development is not the technology to search warehouses faster, but the ability to create entirely new items that are not in the warehouse. Scientists estimate that molecules with

properties that can become drugs theoretically amount to 10 to the 60th power to 10 to the 100th power in number. Since the number of stars in the universe is estimated at about 10 to the 24th power, the size of the possibilities that chemistry holds is immeasurably larger than the stars in the night sky. Such numbers come from the power of combination. Even with just ten types of Lego blocks, the number of ways to connect twenty of them already multiplies beyond what our minds can count.

In fact, according to the Lego company's calculations, the shapes that can emerge from stacking just six completely identical blocks exceed 900 million. However, there are over a hundred types of atoms that make up actual molecules, and if you add the ways of connecting these atoms into chains, branches, and rings, that combination multiplies at a much fiercer speed than Lego. The number of compounds that humans have made and tested in laboratories for a century is incalculably small compared to this vast space.

For a computer to draw new molecules in this vast space, there must first be a way to transcribe the molecules into numbers and letters. A widely used method is SMILES. SMILES is a notation that writes down atoms like carbon, oxygen, and nitrogen, and the bonds between them, with a few alphabets and symbols, writing out the molecule in a single line like a short sentence. Just like taking dictation as a person dictates, the computer learns how to read and write the appearance of molecules by looking at the letters called SMILES. For example, caffeine, which is in coffee and tea, is expressed entirely as a single character string: CN1C=NC2=C1C(=O)N(C(=O)N2C)C.

What order carbon, nitrogen, and oxygen are connected in, and where they form rings, is contained without omission within this short line. The problem lies in the fact that the line of letters is too sensitive. Just as a sentence fails to make sense if even one spelling is wrong, a SMILES string creates a fake molecule that cannot exist chemically if exactly one letter is off. Because such illusions come out mixed in every time a generative model pours out tens of

thousands of candidates, researchers had to wrestle with the weakness of this notation for a long time.

A notation called SELFIES was created aiming directly at this weakness. A character string written in SELFIES mathematically guarantees that no matter how it is combined, it is transferred into a molecule that can actually exist. Like a puzzle cut in advance so the edge shapes do not mismatch each other, SELFIES eliminated the very room for connecting things incorrectly.

As this notation began to be widely used, the process of training artificial intelligence models became noticeably stable, and it became possible to trust each character string and translate it into a molecule. However, the fact that both SMILES and SELFIES are single-line character strings remains the same, so they cannot separately inform us what shape the atoms are placed in within actual space.

Graph neural networks (GNN) are a technology that treats molecules not as one-line character strings but as mesh-like structures themselves. If you place each atom as a dot and draw the chemical bonds between atoms as lines connecting those dots, a molecule takes on a mesh shape similar to a subway map. Graph neural networks read this map as it is and learn both the electronic environment around each atom and its three-dimensional shape. Models called graph convolutional neural networks or graph attention networks, based on this learned information, can predict quite accurately in advance how well a certain molecule will bind to a target protein and how well it will dissolve in water.

Thanks to this technology, the speed and accuracy of virtual screening, where computers first filter tens of thousands to hundreds of thousands of candidate molecules without humans examining each one individually, have greatly improved.

Graph convolutional neural networks work by gathering information about neighboring atoms across multiple layers to summarize the characteristics of the entire molecule, while graph attention networks additionally assign weights themselves to determine which neighboring atoms are more important. An atom placed in the center of the target protein's binding pocket and an atom placed at the edge of the molecule have different impacts on binding strength, and the attention method distinguishes these differences on its own. Thanks to this predictive ability, researchers can filter out most candidates within computers without individually synthesizing and testing millions of candidates in test tubes, putting only a few hundred truly promising ones on the experimental bench.

If you have a language for reading and writing molecules, the next step is to actually create new text in that language. Three technologies responsible for this creation are the Variational Autoencoder (VAE), Generative Adversarial Network (GAN), and diffusion models. A Variational Autoencoder learns numerous compounds and consists of an encoder that compresses the complex shape of each molecule into one coordinate and a decoder that revives that coordinate back into an intact molecule. Within this compressed coordinate space, moving the coordinate just slightly allows you to obtain a molecule with a similar framework to the original molecule while having slightly improved properties like solubility or toxicity.

This method resembles a chef's notes when finding new dishes by gradually changing the proportions of cooking ingredients. Just as reducing salt by only half a teaspoon and adding a bit more sugar produces a dish that is not entirely different but has the same basic flavor structure as the original dish while being much improved. An advanced model called Junction Tree Variational Autoencoder builds molecules using small chunks that are already chemically meaningful as units instead of adding atoms one by one, making the results much more realistic.

Generative Adversarial Networks compete in a different way. It is a structure where a generator that continuously creates fake molecules and a discriminator that distinguishes whether the molecule came from a real drug database or was created by the generator stand against each other and build their capabilities together. The generator, trying to deceive the discriminator, increasingly produces molecules that are difficult to distinguish from real ones. Insilico Medicine, using a model called GENTRL that added reinforcement learning to this Generative Adversarial Network, once found a candidate substance that inhibits DDR1 kinase, an important target in fibrosis and neurodegenerative diseases, in just 21 days.

However, this technology also has clear weaknesses. The training process is mathematically unstable, and the model often falls into a phenomenon called mode collapse, where it loses diversity and repeatedly outputs only a few similar-shaped molecules. Due to this instability, in recent years there is a clear trend in the new drug design field where the focus is shifting from Generative Adversarial Networks to diffusion models.

Diffusion models are a technology that takes hints from the process of restoring photographs. It involves training by gradually adding noise to a clean photo to make it completely blurry, then stripping away the noise layer by layer in reverse to restore the original photo. If you apply this principle directly to molecules, clear molecular structures emerge as you start from a random noise blob and strip away the noise step by step. For a drug to actually work in the body, it must fit into the target protein in three dimensions, not two dimensions, and recent 3D diffusion models have mathematically refined structures that preserve their properties even when molecules are rotated or moved, filtering out unrealistically twisted molecules from the start.

Models like GeoDiff, DiffDock, and DiffSBDD read the shape of the target protein's binding pocket and draw both the atomic arrangement and 3D coordinates that fit precisely into those curves. A model called DiffLinker even

creates a structure that naturally connects separate compound fragments in three-dimensional space. The GlintDM model released in early 2026 simultaneously optimized binding affinity and drug-like properties while significantly reducing computational costs, and the DiffGui model released the same year verified that the designed molecules actually work by conducting wet experiments in actual laboratories where materials are synthesized and tested, not just within computers.

The method of building molecules by directly using information about the target protein's binding pocket is called structure-based generation, and the method of building new molecules with similar properties based only on the characteristics of known active substances without the target's three-dimensional structure is called ligand-based generation. If a protein structure has been experimentally determined, the structure-based method produces more accurate results, but because there are still many targets whose structures are unknown, both methods are used side by side even today.

No matter how many molecules diffusion models or generative adversarial networks pour out, some of them are bound to stick well to the target but be harmful to the body. Reinforcement learning is the technology that guides the direction toward satisfying multiple conditions at the same time.

Reinforcement learning is a learning method in which an agent inside a computer slightly changes the molecular structure, receives rewards and punishments from an evaluation function that scores the result, and learns for itself how to accumulate many points. At this time, the score is not based solely on efficacy. ADMET, an index that encompasses how well it is absorbed, distributed, broken down, and excreted in the body, and whether it has toxicity - that is, absorption, distribution, metabolism, excretion, and toxicity - is reflected in the score, along with the degree to which it dissolves in water, the permeability of the blood-brain barrier that serves as a gateway to the brain, and the burden it places on the heart.

A platform called REINVENT 4, refined by AstraZeneca researchers, combines reinforcement learning with artificial intelligence that handles molecules like a language model, repeatedly refining the molecular structure to increase the effect on the target while lowering the risk of side effects caused by acting on the wrong places. After 2025, research is also continuing to combine reinforcement learning entirely with diffusion models to design molecules that fit multiple targets simultaneously, complete with three-dimensional structures.

Between 2025 and 2026, research combining reinforcement learning completely with other technologies has increased significantly. A method called MDRL weaves together diffusion models and reinforcement learning entirely to design molecules that simultaneously fit multiple different targets, complete with three-dimensional structures. A method called POLO refines candidate substances by reflecting human preferences across multiple rounds as if having a conversation, and a method called MolMem remembers the results of past attempts to help reach the goal with only a small number of attempts. Reinforcement learning itself is rapidly evolving in a direction that allows trial and error to be experienced in advance inside the computer, using those memories for the next design.

This technology for building molecules does not stop at small chemical compounds. In 2025, a model called Chai-2 released by a company called Chai Discovery successfully designed antibodies completely from a blank slate. Instead of copying previously known antibodies or binding substances, it drew new antibodies and nanobodies using only the information of the targeted proteins. Nanobodies are binding proteins that are naturally produced only in the bodies of animals like camels or alpacas, and they are much smaller and easier to handle than normal antibodies. The researchers selected fifty-two targets they had never dealt with before and tested them; even though they only made around twenty candidates per target, they found antibodies that actually stuck to half of them, and it took less than two weeks from starting the

design to confirming the results in the laboratory.

Compared to previous generations of computer design technology, the success rate jumped more than a hundred times. Generative technology, which used to build new small molecules, has now reached out to the design of antibodies, the complex protein weapons that protect the body. More than 86 percent of the antibody candidates completed in this way were even evaluated as having stability and production potential comparable to actual therapeutic antibodies, and in November of the same year, the results of antibody designs against even more difficult-to-handle targets were additionally revealed in follow-up research.

If representation is the foundation, generation is the work of actually moving one's hands on top of that foundation. The work of a human chemist drawing and refining molecules by hand over several months is now drawn and shown by a computer as tens of thousands of alternatives in a day or two. Once the diffusion model and reinforcement learning join hands to build a new molecule that meets the goal, the remaining task is to push that result out into the world.

The stage where these two things happen together - continuing to push the newly built molecule toward a chemical space where no one has gone before, and finding a way to reuse a drug that is already out in the world for a completely different disease - is exactly expansion. Only upon reaching this stage does the imagination of the artificial intelligence that was trapped inside the laboratory finally lead to a drug that will be placed before a patient.

Artificial intelligence that reuses drugs starts from weaving the relationships among diseases, targets, and drugs into a single large knowledge net. Graph neural networks and natural language processing technologies scan millions of medical papers and gene expression data, finding hidden links between drugs and diseases that human researchers have not yet noticed. A single knowledge net like this can hold thousands of diseases, tens of thousands of

drugs, and millions of relationships connecting them. BenevolentAI in the UK analyzed this knowledge net and discovered that baricitinib, a treatment for rheumatoid arthritis, could calm the excessive inflammatory response caused by COVID-19 while simultaneously blocking the path the virus takes to enter cells.

This prediction led to actual large-scale clinical trials and eventually progressed to receiving emergency use authorization as a COVID-19 treatment. This approach of adding a new use to a drug whose safety has already been confirmed is attractive in that it can reach patients in a much shorter time than the path of making a new molecule from scratch.

However, no matter how plausible a molecule the diffusion model and reinforcement learning draw, it is of no use if that drawing cannot be produced in an actual laboratory. Just because atoms and bonds fit together logically on a computer screen does not mean that the molecule can actually be synthesized. It is the same logic that, while any shape of building can be drawn on architectural blueprints, you must reconsider whether materials can be obtained and whether columns can be built to support that weight when actually erecting the building on that land. Among researchers, the term "elephant in the laboratory" is even used to describe this issue. It means a big problem that everyone knows about but rarely manages to handle. The technology that bridges this gap is retrosynthesis planning.

Finding synthesis pathways for complex target molecules by working backward from inexpensive raw materials available on the market is no different from exploring chemical space, and the number of candidate pathways can reach millions. Tools like AiZynthFinder and AskCos analyze these millions of chemical reaction datasets using deep learning and tree search to find inexpensive, high-yielding pathways. This tree search works by trying multiple pathways in advance and selecting routes with high success probability, resembling how reinforcement learning learns by rewarding good

choices.

Making synthesis more feasible is not as simple as it sounds. If molecules are built using only well-known reactions and materials, synthesis becomes easier, but the resulting compounds resemble existing ones and easily lose novelty. Conversely, if one pushes boldly into unfamiliar corners of chemical space, original molecules can be obtained, but whether they can actually be made becomes difficult to guarantee. A 2025 study framed this problem as the challenge of pursuing three rabbits simultaneously: diversity, synthesizability, and biological efficacy. The field's longstanding concern is that pursuing one often means losing another.

In July 2026, Insilicho Medicine announced a new evaluation standard called ChemCensor at the International Conference on Machine Learning (ICML) held in Seoul. Rather than mechanically counting correct answers as the existing method did, this represents an attempt to evaluate the quality of retrosynthetic pathways in a way closer to how actual chemists judge them by looking at reaction centers and functional groups. Recently, research like PrexSyn and SynLlama has also been ongoing, aiming to embed synthesizability from the beginning of the molecule-building process itself, rather than as a final filtering step.

Some laboratories have entrusted this process entirely to robots, experimenting with automated laboratories that run design, synthesis, and testing in sequence without human involvement. Since robots repeat a set sequence on their own without requiring human oversight throughout the night, the time needed to validate a single candidate has noticeably decreased.

As of 2026, there is a tally suggesting that artificial intelligence-derived drug candidate pipelines worldwide reach approximately 170. However, not all prospects are rosy. In 2020, halicin, a substance filtered by artificial intelligence from diabetes treatment candidates, received significant attention

from media and academia as a drug candidate with strong antibiotic properties. A follow-up study published in 2025 reconfirmed that halicin suppressed seventeen out of eighteen bacterial species resistant to multiple antibiotics.

Yet halicin still has not entered human clinical trials. Weaknesses emerged too late: low absorption into the body, too-rapid excretion, and variable efficacy depending on infection conditions. The binding strength that looked perfect on a computer screen does not work before a living body in this way.

Companies that prided themselves on leading artificial intelligence drug development were no exception before this barrier. The United States-based Recursion Pharmaceuticals discontinued development of several candidate molecules throughout 2025. REC-994, targeting cavernous hemangiomas, a cerebrovascular malformation, showed underwhelming efficacy in mid-stage clinical trials and development was halted. REC-2282, targeting neurofibromatosis type 2, took the same path.

The trajectory of this company shows that no matter how superior computational power is, its results do not directly lead to good drugs. In drug development, what artificial intelligence can reduce is the time spent finding and refining candidate substances, not the safety and efficacy thresholds that the human body demands.

There is a common reason these setbacks are repeated. Deep learning models use data from drugs that have already succeeded in synthesis and completed clinical trials as their training material, but this data is concentrated in a very narrow and biased region when viewed against the entire chemical space. The more the model ventures beyond this familiar region to draw unfamiliar molecules it has never seen before, the greater the risk that predictions will diverge from reality.

Researchers call this domain shift, and this is the reason why the procedure of reconfirming binding strength and safety predictions made by computers through experiments and clinical trials cannot be skipped. This is why drug development teams employ a portfolio strategy alongside this approach, conducting experiments on multiple candidates with different properties simultaneously rather than relying on a single choice made by the computer, thereby distributing risk.

This competition is not someone else's business for Korean companies. Hanmi Pharmaceuticals designed HM17321, a candidate substance for obesity treatment, using its own artificial intelligence platform HARP. It works via a pathway different from existing appetite-suppressing drugs, targeting the effect of reducing fat while preserving muscle, and has received approval from the United States Food and Drug Administration for Phase 1 clinical entry. Syntekabio secured multiple biobetter candidate substances redesigned to have superior properties compared to Keytruda, a widely used antibody drug, using its own platform. Standiame has partnered with Hanmi Pharmaceuticals in the search for candidate substances.

Isomorphic Labs, a spinoff of Google DeepMind, also unveiled an engine called IsoDDE in February 2026 based on AlphaFold 3, introducing a function capable of drawing antibodies from scratch. The first human clinical trial is scheduled to begin within 2026. The company attracted over 600 million dollars in new investment in 2025 alone. This engine demonstrated accuracy more than twice that of AlphaFold 3 on Runs N' Poses, a newly created binding prediction benchmark, and even predicted previously unknown binding sites on the protein cerebellon using only sequence information, with this prediction later confirmed by subsequent actual research.

The competition to draw new drug candidates from the ground up has now become a phenomenon occurring simultaneously worldwide, no longer just a story among a few leading companies.

The hands that build are getting faster and faster. If it was considered a surprise in 2019 when GENTRL found a DDR1 inhibitor candidate in 21 days, ISM001-055 in 2026 showed that a substance created by those fingertips could advance all the way to Phase 3 clinical trials. However, the bottleneck has not disappeared, but merely moved elsewhere. In the past, the very act of imagining a usable molecule was difficult. Now, the bottleneck remains in confirming whether the imagined molecule can actually be created, and whether that molecule is truly safe and effective inside a living body. With the three pillars of representation, generation, and expansion in place, the speed of drawing up new drug candidates has already surpassed human hands. The remaining battle is to place that drawing in front of real-world chemistry and biology and make it endure to the end.

Just as a single drawing that started in a laboratory in Shanghai in 2022 has reached the threshold of Phase 3 clinical trials in 2026, it is highly likely that we will face a moment in the next few years when a drug drawn by a computer from the very beginning reaches pharmacy shelves. New drug design, which used to rely on the intuition and fingertip senses of chemists, is thus shifting its center of gravity to a process shaped together by data and algorithms. This change is connected to a trend that is rewriting the research methods of the entire pharmaceutical industry.

Although the hands that draw candidate substances have become faster, the hands that verify those drawings are not yet as fast. An industry analysis published in 2026 revealed that the bottleneck in new drug development has shifted from the stage of finding candidate substances to the stage of laboratory verification. Eli Lilly in the United States has already been operating a robot laboratory with a company called Strateos since 2020, which bundles design, synthesis, purification, and analysis into one and runs with almost no human intervention. The Korea Health Industry Development Institute has also pointed to equipping autonomous laboratories - where robots and artificial intelligence continuously handle design, synthesis, and experiments without

human hands - as a task to be solved in the future.

There is a little-known situation behind this trend. For artificial intelligence to reduce trial and error, it must learn not only what worked, but also what did not work and why. However, records of failed compounds are rarely left in academic papers or public databases. Companies quietly put failed candidates in a drawer and do not announce the reasons for that failure to the world. Artificial intelligence that has learned only by collecting successful cases often fails to read the signs in advance, even when failure is right in front of it. It seems that this lack of data will not be easily filled, no matter how many autonomous laboratories are built.

Generative models do not only build molecules that stick well to targets. They are also becoming increasingly skilled at building molecules that avoid the claims of already registered patents. Recently, methods have even emerged that embed this goal into the score calculation of reinforcement learning itself. It is a method that gives penalty points to molecules with structures close to existing patented compounds, and gives rewards to molecules that find empty spaces where no one has yet claimed rights. The ones using this tool are not only the companies that first created the new drugs. Companies that make generic drugs are also designing similar substances that avoid patents using the same technology, hastening disputes against the patents of the original companies. A US court ruled in 2022 that artificial intelligence cannot be a patent inventor.

However, when the one that drew the molecule is an artificial intelligence and a human only refined the synthesis pathway, there is still no clear answer as to whose name should be written in the inventor box. The patent system is a promise that grants monopoly rights for a certain period in exchange for disclosing a new invention to the world. If the speed at which machines find and navigate loopholes in claims is much faster than the speed at which humans revise laws, it is highly likely that the weight of this promise will

become increasingly light. Even if it has the effect of bringing forward cheap generic drugs, the problem where this trend clashes with the original intent of patents - which is to protect inventions and allow research costs to be recovered - is expected to remain noisy in the future.

2.3

In Silico ADMET and Toxicity Prediction System

In the pharmacies of the United States in the 1990s, the antihistamine terfenadine, sold under the brand name Seldane, was widely prescribed to people suffering from seasonal allergies. It was highly effective at calming sneezes and runny noses, but accidents continued to occur where some patients who took this medication collapsed from an irregular heartbeat when they consumed it together with certain antibiotics, antifungals, or even grapefruit juice. The cause was an arrhythmia, a condition where the electrical signals of the heartbeat are disrupted because the drug blocked the potassium channels of heart cells, a door commonly called the hERG channel.

At the time of development, no one anticipated that this drug would fail to break down in the liver and would accumulate in the blood to threaten the heart. Eventually, this medication quietly disappeared from the market, and the pharmaceutical industry spent a long time searching for a way to find out in advance, before clinical trials, how a single new drug moves inside the human body and where it becomes dangerous. Artificial intelligence in the late 2020s has grown into an in silico prediction system that provides the answer to this very question, previewing the fate of a candidate substance inside a computer before it even enters the human body.

It takes an average of more than 10 years and hundreds of billions of won for a single new drug to make its way from a laboratory compound to a pharmacy shelf. However, out of the candidate substances that enter clinical trials, the proportion that survives to final approval is less than one in ten. According to a study that analyzed cases of clinical trial failures from 2010 to 2017, 40 to 50 percent of the reasons for failure were because the drug's effectiveness did not appear as expected, nearly 30 percent were due to toxicity that the body could

not handle, and 10 to 15 percent were because the way the drug was absorbed and disappeared in the body was poorly designed in the first place. If you add the toxicity and pharmacokinetics issues together, nearly half of clinical failures occur in the area of the five letters of absorption, distribution, metabolism, excretion, and toxicity, or ADMET.

The United States National Institutes of Health has stated that 22 percent of new drug research and development costs go directly into identifying and solving this ADMET problem. In the pharmaceutical industry, the paradoxical trend of new drug development costs actually increasing as time goes by is called Eroom's Law, flipping Moore's Law, which states that semiconductor performance steadily improves, backward. If the toxicity of a candidate substance is only revealed when it reaches animal testing and clinical trials involving humans, the time and money invested up to that point disappear entirely.

For a long time, pharmaceutical companies have relied on the method of feeding candidate substances to animals like mice and monkeys and observing their reactions in order to filter out these risks in advance. The problem is that animal bodies and human bodies differ quite a bit in their genes and physiological structures. Cases repeated where drugs confirmed as safe in animal tests caused unexpected side effects in actual clinical trials involving humans, and the academic community suffered headaches for a long time, calling this gap the translational research gap. Through the Food and Drug Administration Modernization Act 2.0 in 2022, the United States Food and Drug Administration legally removed the 84-year-old regulation that animal testing must be mandatorily included in new drug approval applications.

Going one step further in April 2025, they released a roadmap to make animal testing an exceptional procedure within the next 3 to 5 years, starting with antibody treatments, and announced that they would fill that empty space with new evaluation methods like artificial intelligence-based toxicity

prediction models, organoids, and organs-on-chips. The Food and Drug Administration Modernization Act 3.0 bill, which nails down the complete reflection of this law in actual review regulations, also passed the Senate in December 2025 and the House of Representatives in July 2026 sequentially, and as of August 2026, only the procedure of passing over to the President's signature after reconsideration by the Senate remains.

So-called Early Filtering, which pre-filters problematic candidate substances before entering the expensive gateway of clinical trials, has taken over the job that animals in laboratory glass bottles used to do, becoming the first gateway of new drug development.

The five letters of ADMET are themselves a story of the journey a drug experiences inside the human body. The first letter, A, is absorption, which refers to the process of a pill passing through the stomach and intestines and seeping into the blood. D is distribution, which is the process of the drug traveling throughout the whole body along the blood and then staying in places like the liver, brain, or muscles. M is metabolism, which means the process of the liver chopping the drug into small pieces and changing it into other substances, much like a disassembly line in a factory.

E is excretion, which is the process of leftover drug residues going outside the body through urine or feces after passing through the kidneys. The final T is toxicity, which points to whether the drug injures organs like the liver, heart, or kidneys during all of this journey. If a problem occurs in even one of these five stages, no matter how effective a candidate substance is, it ultimately fails to be born as a medication.

The academic discipline that guesses in advance how a molecule will move inside the body just by looking at its single structure is called the Quantitative Structure-Activity Relationship, or QSAR for short. It is a theory that started from the idea that there are certain rules between the appearance and the properties of a molecule. Early QSAR stayed at the level of statistical formulas

that gauged drug-likeness by weighing just a few numbers like molecular weight or fat solubility, such as Lipinski's Rule of Five. However, the chemical space of compounds that must be reviewed as new drug candidates today has grown wide enough to reach millions or tens of millions, and this vast space can no longer be handled with old linear formulas alone. Therefore, researchers began to combine QSAR with Graph Neural Networks, or GNN for short, which recognize a molecule as a net made of dots called atoms and lines called chemical bonds, along with large language models and transformer architectures that write down a molecule as a character string called SMILES and read its grammar.

ADMET-AI, which came out as a result, scans the absorption and toxicity of millions of compounds in just a few hours, and Deep-PK produces up to seventy-three pharmacokinetic, toxicological, and physicochemical indicators all at once using graph-based molecular signatures. These models have grown by taking vast public databases that anyone can download, such as ChEMBL, DrugBank, and Tox21, as their learning materials.

The international guideline ICH M7, which judges whether impurities that can be mixed into pharmaceuticals have the risk of damaging genes and causing cancer, is a representative example of recognizing QSAR predictive values as actual regulatory review data. This guideline stipulates that two predictive models with different principles - namely, a model that translates experts' chemical knowledge into rules and a model that learns statistically - are run side by side, and if different conclusions arise, a person must directly make a judgment. In 2025, AmesNet appeared, which increased sensitivity by nearly 30 percent compared to existing machine learning techniques even for unfamiliar chemical structures, raising the prediction of mutagenicity of novel skeleton compounds, which had long been a stumbling block, to the next level.

Underlying what these models learn are the ToxCast and Tox21 databases operated by the United States Environmental Protection Agency, and the

latest edition released in August 2025 contains in vitro reaction information for nearly ten thousand chemicals, making it one of the leading public assets for training toxicity prediction artificial intelligence. Pharmaceutical companies are using such public data in training alongside their own compound libraries to filter out impurities suspected of mutagenicity or carcinogenicity right from the design stage before synthesis.

Going beyond predicting the properties of candidate substances, there are also active attempts to newly draw molecules with good properties from scratch inside a computer. The Retro Drug Design framework trims the molecular structure on its own in a direction that simultaneously satisfies six ADMET conditions that are easy to conflict with each other, such as drug efficacy, bioavailability, metabolic stability, and low toxicity, and indeed, the kinase inhibitor candidates designed in this way showed a high success rate in experimental verification. ADMETrix, which weaves the generative artificial intelligence REINVENT and ADMET-AI into one, calculates several ADMET indicators together immediately as soon as a new molecule is drawn, allowing real-time switching between design and evaluation.

DrugEx v2 brings Pareto-style reinforcement learning, which finds a balance point without giving up any one when multiple goals collide with each other, into drug design, picking out molecules that stick well to target proteins while avoiding the wrong targets. Just as a cook weighs taste, nutrition, and cooking time all at once when making a recipe, these models are serving multiple plates of efficacy and safety at the same time.

Figuring out the properties of each individual molecule and drawing how that molecule goes up and down over time inside an actual human body are problems on a different level. The technology that bridges this gap is physiologically based pharmacokinetic modeling, or PBPK modeling for short. It is easy to understand if you view it as a method that places the human body as a map made up of several cities like the heart, liver, kidneys, and muscles,

and uses blood flow as the roads connecting those cities, calculating how long a car called a drug stays in which city over time. Artificial intelligence analyzes the SMILES structure or fingerprint information of a molecule and instantly extracts core figures such as partition coefficient, aqueous solubility, degradation rate in liver microsomes, and the ratio of binding to plasma proteins, and when these figures are put into the PBPK simulation engine, a curve is drawn showing how the blood concentration goes up and down over time after intravenous injection or oral administration.

In a validation study that combined chemical language models or graph convolutional neural networks with PBPK, the predicted values of indicators such as maximum blood concentration, its peak time, area under the curve, and half-life showed a correlation reaching a coefficient of determination of 0.61 to 0.93 with actual values, and the absolute average fold error, which indicates the size of the error, also stayed between 1.84 and 2.8, showing precision comparable to existing animal testing extrapolation methods. In the United States, out of 245 new drugs approved between 2020 and 2024, 65 cases, or 26.5 percent by ratio, submitted a PBPK model as a core basis for review, and among them, cases used for drug-drug interaction evaluation stood out at 81.9 percent, and it was exceptionally actively used in the field of anticancer drugs at 42 percent.

Such precision does not stop at setting a safe starting dose in clinical trials where a new drug is first administered to people, but is being used as a foundation for patient-customized precision medicine that calculates in advance the drug exposure of groups that are difficult to directly participate in clinical trials, such as the elderly with reduced liver function or children whose weight and organ sizes differ from adults. The United States Food and Drug Administration has separately prepared guidelines that officially recognize the role of PBPK models in drug interaction evaluation in line with this trend.

Predicting cardiotoxicity is considered an urgent task to be handled in the field of new drug development. As seen in the case of terfenadine, if a drug blocks the hERG channel, which is the potassium channel of heart cells, the electrical signal of the heartbeat is disrupted, which can lead to life-threatening arrhythmias. Cisapride, which helped gastrointestinal motility, also withdrew from the market for similar reasons, and it was a case that showed the fact that hERG-related arrhythmias are a risk that appears repeatedly across several treatment fields without stopping at just one class of antihistamines. An artificial intelligence-based quantitative continuous scoring system created by Mamoshina researchers learned vast amounts of compound information contained in DrugBank and medical dictionary databases, sorting out potential cardiotoxicity in advance at a level of 79 percent area under the curve in validation data.

In 2025, hERG-MFFGNN, a graph neural network model that combines multiple features all at once, appeared in a way that even explains the basis of judgment, and a dataset called DICTrank, which standardized and gathered reliable cardiotoxicity data, was also announced together. In the same year, an XGBoost consensus model that combined balanced training data and various variables received a report card with a sensitivity of 0.83 and a specificity of 0.90, and with the emergence of even hERG-LTN that brought in the embedding technology of large language models, prediction methods are becoming more diverse with each passing year.

In a paper that collected and reviewed 25 related studies, several models with an area under the curve exceeding 0.70 or an accuracy above 0.75 in external validation were identified, receiving the evaluation that they have taken a big step closer to a level usable in the actual field.

Hepatotoxicity, and specifically drug-induced liver injury, or DILI, is particularly difficult to predict because the causes of onset cannot be neatly separated into a single factor. This is because the amount and types of liver enzymes differ

from person to person, and even the same medication can leave one person's liver intact while damaging another's. Common polymorphisms in cytochrome family enzyme genes - such as differences in the activity of CYP2D6 and CYP2C19 that people are born with - mean that even the same drug dose becomes excessive for some people and insufficient for others.

Recently, there have been ongoing attempts to learn this genotypic information together with compound structure information to identify hepatotoxicity risks that stand out only in specific genotypic populations. Recent artificial intelligence systems have achieved results of identifying drug-induced liver injury with up to 89 percent accuracy by simultaneously examining the physicochemical properties of drugs along with multi-omics data encompassing genes, proteins, and metabolites, as well as off-target interaction networks.

In early 2026, a multimodal deep learning method that combines information extracted by different language models such as ChemBERTa, RoBERTa, and SMILES2Vec was published in academic journals, making much finer-grained prediction possible than the previous approach of examining only molecular structure. There was also a symbolic event. In 2025, the U.S. Food and Drug Administration accepted an artificial intelligence-based digital liver model submitted by Absentia Labs as a certification target for the IStand program, which reviews innovative new drug development approaches, and this was the first instance of a prediction tool created by artificial intelligence entering this certification process.

Although it has only passed the first gate of a three-stage certification process, it remained a meaningful event in that a virtual liver inside a computer has begun to fill one section of new drug review documents in place of a real human liver.

Drugs do not quietly bind only to their originally intended target protein. Just as a single key fits partially into multiple locks, when a new drug candidate

gets entangled with other receptors or enzymes that were not originally targeted, causing unexpected side effects, this is called an off-target effect. The Liu research team used a multitask graph neural network where a single model processes multiple prediction tasks simultaneously to present a method that precisely captures the off-target interactions of compounds, and demonstrated that this method can filter out serious risks such as drug-induced kidney injury at the compound optimization stage.

The ProTox family of web servers that delve into the relationship between structure and toxicity has also grown rapidly. The latest version, ProTox 3.0, calculates toxicity indicators reaching sixty-one kinds, including median lethal dose, carcinogenicity, mutagenicity, and immunotoxicity, by combining molecular similarity with machine learning. Before entering costly animal toxicity testing, it has become possible to filter out compounds with serious defects from candidate groups using just this one free web tool.

As the elderly population grows and more patients suffer from chronic diseases, it has become common for one person to take multiple medications at once, and identifying drug-drug interactions, or DDI, in advance has emerged as a problem directly linked to patients' lives. Just as when terfenadine meets antibiotics or grapefruit juice, when one drug blocks or enhances liver enzymes that break down another drug, the concentration of the second drug rises far higher or drops far lower than expected. In the past, confirming such interactions remained at the level of searching existing literature one by one, but now artificial intelligence that reads vast medical literature, electronic health records, and three-dimensional structures of compounds simultaneously has taken its place.

Large language models such as BERT-DDI identify drug-to-drug risk signals buried in millions of clinical records and adverse event reports at the sentence level. The concept of DTiGEMS+, which drew the relationship between drugs and target proteins as a graph and inferred new combinations on the

relationship network, led to subsequent knowledge graph neural networks like SumGNN, reaching a level where it can identify risks of new drug combinations that have never been reported in clinical settings. The prediction results obtained this way are already connected to clinical decision support systems that display warnings on the screen the moment a doctor writes a prescription, and they are becoming not just calculations that end inside a computer, but safety devices that actually work in the clinic.

Even after a drug comes to market, artificial intelligence continually reviews actual clinical data such as electronic health records and insurance claim data, and post-market drug surveillance work that identifies risk signals of drug combinations that did not emerge during the clinical trial phase, however late, is also being carried out.

When a molecule's SMILES string or three-dimensional structure, along with physicochemical properties such as melting point or lipophilicity, is entered, the output on the other side includes ADMET property predictions along with the applicability domain, which indicates how far this prediction can be trusted. The applicability domain is a measure indicating how similar the molecule being evaluated is to the compounds the model has learned, and the Organization for Economic Cooperation and Development recommends that this domain must be clearly specified when validating QSAR models. Just as when you turn onto an unfamiliar road not on a map and the navigation guidance suddenly starts going haywire, when encountering a new chemical structure far removed from the training data, even the most sophisticated model's predictions can be greatly shaken.

ADMETlab 3.0, one of the widely used hybrid platforms today, gathers information on over 400,000 compounds from public databases like ChEMBL, PubChem, and OCHEM, and uses SMILES as a standard input to calculate 119 properties all at once. It is praised for catching both speed and accuracy by combining traditional molecular descriptors alongside a multi-task directed

message passing neural network that handles multiple prediction tasks at the same time. However, a critical review paper published in 2026 pointed out that only a handful of top-ranking models succeeded in reproducing all 22 standard evaluation items on the ADMET leaderboard of the representative benchmark repository, Therapeutics Data Commons (TDC). It noted that some models did not disclose their code or even had flaws where training data and test data were mixed. This left a reason to double-check before simply believing the impressive performance numbers.

The loop of exchanging predictions and verifications is also getting tighter. The so-called design, make, test, and analyze cycle is spreading to various research institutions. In this cycle, a robotic lab immediately synthesizes the candidate substances that AI has picked as promising, automatically runs test tube experiments, and then feeds the results back into the prediction model. Unlike the old way where a person ordered compounds one by one and scheduled experiments, this cycle runs multiple times a day and continuously refines the prediction model itself.

As ADMET prediction becomes more precise, the speed of this cycle also gets faster, and the new experimental data piled up in this way inversely hones the next-generation prediction models more sharply.

The threshold that this seemingly fast and precise AI prediction must cross to enter the clinical and regulatory fields is still high. It is not rare for an excellent model that hovered around a coefficient of determination of 0.8 in internal testing to see its external validation performance plunge to the 0.3 range when it meets a compound with a new scaffold that was not in existing databases. It is also a hurdle that most of the toxicity and pharmacokinetics data piled up so far leans heavily toward specific drug groups that were already easy to synthesize and successfully commercialized, or toward Western populations. Models trained on this data carry the risk of giving wrong answers when evaluating unique metabolic traits of minority races or new treatment

methods like cell and gene therapies. If metabolic enzyme variations that stand out only in specific populations are not fully reflected in the training data, the model might miss side effects that appear unusually large only in that group.

In many cases, they remain in a black box state where a deep learning model tangled with tens of millions of parameters cannot explain in words humans understand why it pointed out a specific compound as hepatotoxic or what structural features cause arrhythmia. Because this lack of transparency is hard to accept as it is during the review process that deals with patients' lives, the prediction values put out by AI must be checked one more time through physical biological systems like organoids or organ-on-a-chip before they are finally recognized as doing their proper job.

In January 2025, the US Food and Drug Administration released a draft guidance covering how AI should be used in regulatory decisions for drugs and biological products, and presented a seven-step procedure to verify reliability by dividing the risk level according to the situation where the model is used. This guidance broadly covers the stages where AI steps in, from nonclinical and clinical to post-market surveillance and manufacturing stages, but it drew a line saying it does not regulate business efficiency improvements that do not affect safety or quality. In September 2024, the European Medicines Agency released a reflection paper on the application of AI across the entire drug lifecycle. It strongly recommended traceability and transparency while pulling AI into existing good clinical practice and quality control principles.

Starting in 2026, the Ministry of Food and Drug Safety of the Republic of Korea also began to gradually bring in AI for summarizing approval data, professional translation, and drafting review reports. Using this as a stepping stone, it set a goal to finish new drug approval reviews within 240 days, while announcing that it will train 1,100 regulatory science experts focused on the AI new drug and bio fields over the next 5 years. Under the awareness that if each country's

regulatory agency keeps its different AI review standards as they are, clinical trial data conducted together in multiple countries could be evaluated differently from country to country, they also started discussions to align common standards through councils like the International Council for Harmonisation (ICH).

From a single prediction value put out by a computer to leading to an actual prescription, these regulations and procedures now dictate the speed just as much as technology.

No matter how excellent a model is, it cannot use its strength if it lacks data to learn from. However, hospitals and pharmaceutical companies consider the clinical records and compound data they worked hard to gather as core trade secrets and rarely let them out. On top of this, as regulations protecting patient information like the Health Insurance Portability and Accountability Act in the US or the General Data Protection Regulation in Europe overlap, a so-called data silo phenomenon has solidified, where data is scattered and trapped inside various organizations. Federated learning is receiving attention as a way to gather and use data from different institutions without tearing down these walls.

It is a method where the original data is left exactly as it is inside the secure servers of hospitals or companies, and only the weights of the AI model that learned from that data are encrypted and sent to a central server to be combined. In the Republic of Korea, the Ministry of Science and ICT and the Ministry of Health and Welfare are spending 34.8 billion KRW together to push forward the K-MELLODDY project from April 2024 to December 2028. The work of honing a model that predicts ADMET and pharmacokinetics together by tying laboratory data and clinical data held by over 20 domestic institutions through federated learning was expanded into a full-scale new project starting in 2025.

The MELLODDY project, conducted by multiple pharmaceutical companies in Europe, provided the foundation, and both projects demonstrated that predictive models could be developed jointly without compromising patient privacy or trade secrets. One domestic pharmaceutical company announced that, in collaboration with Insilico Medicine, it had identified a promising drug candidate compound possessing all five essential criteria - absorption, distribution, metabolism, excretion, and toxicity - in just four months of cooperation, a task that would previously have taken years to complete.

This trend is not confined to the narrow field of ADMET prediction, but is expanding into a larger convergence ecosystem that integrates multi-omics analysis reading both genomic and proteomic information, physical validation devices such as organoids and organ-on-a-chip systems, and electronic health records from actual clinical practice. When a structure is established in which computer predictions, laboratory validation, and hospital prescription records for a single drug candidate compound exchange data with one another, the accuracy of ADMET prediction rises not from the advancement of any single technology, but from the pace at which this entire ecosystem grows and strengthens together.

Within Korea as well, as cases of cooperation between government-funded K-MELLODDY and private enterprises accumulate, this convergence is transforming from a tool available only to a handful of large pharmaceutical companies into shared public infrastructure that multiple institutions can access together. The next deciding factor for ADMET prediction technology is not which company possesses a more sophisticated algorithm, but rather how densely and diversely a company can weave together data and validation methods.

Even at this very moment, servers around the world are calculating the ADMET properties of millions of candidate compounds throughout the night, and compounds showing clear danger signals are quietly removed from the

candidate list before encountering even a single animal or patient. This is a scene that would have been difficult to imagine in the era when terfenadine was sitting on pharmacy shelves. Of course, we cannot promise that current artificial intelligence filters out all risks perfectly. Before new structures it has not learned, it still wavers, and it often cannot explain why it made certain judgments; ultimately, it must undergo physical validation such as organoids and organ-on-a-chip systems before its values are accepted as trustworthy.

Nevertheless, the mesh of early filtering that screens out dangerous candidate compounds at the stage before crossing the clinical trial threshold is becoming finer with each passing year, and each thread weaving that mesh is connected to QSAR and PBPK, graph neural networks and large language models, and federated learning that crosses borders. The procedure requiring drug candidate compounds to first pass through computer simulation as a gateway before touching the human body is solidifying into a standard phase of drug development, transcending pilot projects at a few laboratories.

In September 2025, a patient receiving nexiguan ziclumeran (nex-z), being developed by gene editing therapy company Intellia Therapeutics, experienced grade 4, that is, life-threatening liver enzyme elevation. The patient received treatment at a hospital. The drug was aimed at transthyretin amyloidosis, a genetic disease in which amyloid proteins accumulate in the heart and nervous system, and was the centerpiece of phase 3 clinical trials MAGNITUDE and MAGNITUDE-2, in which the drug enters the body directly and rewrites genes. On October 29, the U.S. Food and Drug Administration issued a clinical hold order for both trials in their entirety, and in November, an octogenarian male patient died following drug administration, compounding the crisis.

Investigation revealed that this death was due to infection rather than liver damage, but the fact that the entire late-stage clinical trial came to a halt in an instant remained unchanged. This therapy's mechanism was not the chemical

structure of a compound, but the very operation of gene scissors within the body. It was a target that existing predictive models trained on SMILES strings and molecular fingerprints had never encountered in the first place. After adding safeguards that monitored liver markers more frequently and screened out at-risk patients in advance, the two clinical trials each resumed their progress only after entering 2026.

As completely new therapeutic approaches like cell and gene therapy increase, a paradox appears to be emerging: the scope of applicability of predictive models based on molecular structure is actually narrowing. Artificial intelligence that reads the toxicity of compounds stands on decades of small-molecule data. However, data that can capture the immune responses and off-target editing caused by gene editing or cell therapy have not yet accumulated sufficiently anywhere in the world. For the time being, even the most sophisticated predictive models cannot replace the surveillance system of the clinical trial site that watches patients directly and captures abnormal signals immediately - this possibility appears quite large.

Efforts to broaden the scope of prediction have also begun anew in Korea. On December 3, 2025, the Ministry of Food and Drug Safety concluded an agreement with Seoul National University's Graduate School of Data Science, and decided to jointly establish an artificial intelligence platform that weaves together clinical data from domestic medical institutions and multi-omics information to map drug responses that differ from person to person. Unlike K-MELLODDY, which filters promising drug candidates at the laboratory stage in advance, this agreement represents an attempt to transfer into data even what happens after a drug enters a patient's body.

The two institutions committed not only to joint research, but also to holding workshops together to manage data quality and develop analytical personnel. As these two initiatives accumulate in parallel, the temporal scope handled by predictive models is expanding as well.

PART 3

Next-Generation Modalities and Smart Drug Delivery

3.1

Biopharmaceuticals and Multi-Target Therapeutics Design (ADC)

On June 22, 2026, for four days, the San Diego Convention Center in California, USA, was so crowded with pharmaceutical and biotech company representatives from around the world that there was barely room to stand. At this event, the world's largest bioindustry trade show, commonly known as BIO USA or Bio International Convention, the Korean company Celltrion set up a booth and conducted over 180 business meetings over four days. More than 2,000 visitors passed through the booth. What Celltrion showcased was not biosimilars, the generic versions of patented biopharmaceuticals that had been their main focus until then, but rather AI-based novel drug development, antibody-drug conjugates, and multi-specific antibody technology that targets multiple targets simultaneously.

It was evidence that the center of gravity in the pharmaceutical industry was shifting from an era when small-molecule compounds were designed one pill at a time by computer to an era when large protein molecules are being completely redrawn using artificial intelligence.

Just a few years ago, designing large pharmaceutical products such as antibodies and cell therapies was arduous work that depended on the intuition of researchers and repeated experimentation. Once a candidate antibody was made and placed in an experimental dish to observe its reaction, if it failed, one had to start from scratch. This trial-and-error process had to be repeated hundreds or thousands of times. In that process, it was not uncommon for more than ten years to pass and costs approaching trillions to accumulate. However, as technology developed that allowed researchers to pre-draw and test the amino acid sequence and three-dimensional structure of

proteins down to the atomic level on a computer screen, the story changed.

At the center of this are large language models and protein language models, along with multi-omics analysis that reads multiple layers of biological data all at once. Design work that used to require spending years at the laboratory bench is now increasingly being completed in days on a server.

An antibody is a Y-shaped protein created by our body's immune system. It recognizes foreign substances that have invaded the body, such as bacteria or viruses, and attaches tightly to them to mark them. Like a key that fits a particular lock, each individual antibody is structured to bind only to a designated target. Our body can produce millions of different antibodies, and selecting one antibody that attaches to only one specific target from among them and mass-producing it for use as medicine is what is called a monoclonal antibody (mAb). From treatments for rheumatoid arthritis to various cancer treatments, most of the antibody medications currently used in hospitals are made this way.

Going one step further from an antibody attaching to only one target, antibodies designed to simultaneously attach to two or more different targets are increasingly appearing. These antibodies are called multi-specific antibodies or bispecific antibodies. With one arm, they grasp a specific protein on the surface of cancer cells, and with the other arm, they grasp T cells patrolling the body, bringing the T cells right next to the cancer cells in the same way a matchmaker would hold each of two people's hands and make them face each other.

The technology that Celltrion unveiled at BIO USA 2026 was an artificial intelligence tool for designing such multi-specific antibodies and pre-evaluating their development potential. As the number of targets increases from one to two, three, and more, the number of combinations that antibodies must recognize increases exponentially, so the artificial intelligence searches the design space that would have been difficult to handle using the

old method where humans laboriously narrowed down candidates by hand.

An antibody-drug conjugate (ADC) is a precision-guided medicine in which a powerful toxic substance is attached to this antibody. It is easy to understand if you think of a delivery package. Inside the box is a chemical substance toxic enough to kill cancer cells - that is, the payload. If you simply throw this box at any house, it would be a disaster, so you need a delivery person who knows the correct address. That delivery person is the antibody.

The antibody recognizes a specific protein that exists only on the surface of cancer cells and delivers the package only to that house. And the linker is the rope that keeps the package tightly bound so that it absolutely does not open during delivery, but loosens only when it reaches the correct house. Only when the three components - the antibody, the linker, and the payload - work in concert is a precision bomb created that explodes the toxic substance only inside cancer cells.

Existing anticancer chemotherapy has operated in quite a different way. It was a method of sending toxic drugs through blood vessels throughout the body to attack rapidly dividing cells indiscriminately, and in that process, the side effects of hair loss, intestinal mucosal erosion, and dropping white blood cell counts were unavoidable. It was close to carpet bombing. Antibody-drug conjugates, by contrast, are more like precision-guided missiles that strike with coordinates precisely marked and fall only at that location.

Even when using the same toxic substance, because they act specifically on cancer cells without touching normal cells, they can produce much stronger effects with much lower doses. The difficulty of actually implementing this precision has long been a bottleneck in antibody-drug conjugate development.

Even before this AI-based design workflow became full-fledged, antibody-drug conjugates had already proven their worth in clinical settings as a modality.

Drugs such as Enhertu, used in breast and gastric cancer treatment, and Trodelvy, used in breast and bladder cancer treatment, have built trust in this field by opening new options for patients for whom existing anticancer drugs had no more to offer. Artificial intelligence is now taking on the role of advancing this already-proven modality to the next generation, faster and with greater precision.

Right here is where artificial intelligence is changing the landscape of antibody-drug conjugate design. Instead of adjusting the antibody's binding site, the linker's chemical structure, and the payload's toxic potency one by one separately, researchers have begun using a reinforcement learning method that adjusts all three factors simultaneously. When an AI model presents a candidate combination, the computer calculates how stably that combination holds up in the blood, how precisely it breaks apart when it reaches a cancer cell, and how much it leaks into normal tissue.

Based on those results, it assigns a score, and if the score is low, it changes the structure and repeats the calculation process tens of thousands of times. In this way, a virtuous cycle is created that simultaneously pursues the dual goals of raising the antibody's target binding affinity and lowering systemic toxicity. In a sense, it is a structure in which the antibody, linker, and payload evolve together, making up for each other's weaknesses.

This cycle created by the antibody, linker, and payload can practically be called a multi-target optimization loop. Looking more closely at this cyclic structure, the reward signals given to the artificial intelligence are largely divided into two types. One is a binding metric that indicates how strongly and for how long the antibody sticks to the surface antigens of cancer cells, and the other is a selective toxicity metric that shows how quietly the payload remains in normal cells and how fiercely it operates only inside cancer cells.

Finding a combination that raises both metrics simultaneously is a multidimensional optimization problem with too many variables for humans

to calculate by hand. While simultaneously testing numerous combinations of antibodies, linkers, and payloads, artificial intelligence searches for combinations that satisfy both metrics together, and it passes only the resulting candidates onto the actual laboratory verification stage. Multiple combinations that were rarely seen in the old method, where chemists and immunologists worked separately, are being discovered through this process.

A technique called quantitative systems pharmacology is also used in this virtual calculation. This is a method of mapping out in advance, through a mathematical model, the entire process of a drug being absorbed, spreading, and breaking down after it enters the body. If the payload's cell-killing performance and the linker's stability are filtered in advance using this model, a significant number of combinations destined to fail can be eliminated before entering actual animal testing and clinical trials. Movements to acquire companies with these design capabilities heated up the pharmaceutical industry in 2025 and 2026. In April 2026, Gilead Sciences agreed to acquire Tubulis, a German company specializing in antibody-drug conjugates, for a total of 5 billion dollars, adding up to 1.85 billion dollars in phased milestones to a 3.15 billion dollar down payment.

Eli Lilly also joined this trend by purchasing Crossbridge Bio, an antibody-drug conjugate startup. Both deals were acquisitions aimed at the artificial intelligence platform itself, which designs linkers, payloads, and conjugation chemistry, rather than targeting a specific single anticancer candidate held by the companies.

The presence of Chinese biotechnology companies in this market is overwhelming. It is even assessed that among antibody-drug conjugate technology export contracts worldwide, candidates originating from China account for about nine out of ninety cases. Among Korean companies, LigaChem Biosciences, formerly known as LegoChem Biosciences, is leading the pack. Armed with its independently developed proprietary linker and

payload technologies, this company exported its technology to Amgen for about 1.6 trillion won, to Janssen for about 2.2 trillion won, and to Japan's Ono Pharmaceutical for over 940 billion won, and just the disclosed contracts combine to exceed 9.3 trillion won.

There are also estimates that the cumulative technology export volume will exceed 10 trillion won when undisclosed contracts are added. LNCB74, a B7-H4 targeting antibody-drug conjugate being developed by LigaChem Biosciences together with the antibody developer NextCure, is moving on to the next stage by applying to the U.S. Food and Drug Administration to add a high-dose cohort in its global Phase 1 clinical trial. The market research firm EvaluatePharma forecasts that the total size of the antibody-drug conjugate market will grow from around 15 billion dollars in 2025 to 57 billion dollars in 2032.

Even at the stage of designing the antibody itself, artificial intelligence has already entered practical use. After learning from vast amounts of antibody sequence data, deep learning-based protein language models create entirely new antibody sequences from scratch that bind strongly to specific targets. Absci, a U.S. artificial intelligence drug discovery company, advanced ABS-101, an antibody designed with its generative artificial intelligence platform, into a Phase 1 clinical trial as a candidate for treating inflammatory bowel disease.

Designed to bind to a target protein called TL1A, this antibody drew industry attention as the first biologic to be administered to an actual person after being drawn entirely by artificial intelligence. It was designed to require a subcutaneous injection only once every four months, and it also focused on lowering the risk of immunogenicity. Initial safety, pharmacokinetics, and immunogenicity data were scheduled to be released in the second half of 2025.

Massive investment funds are also flocking to artificial intelligence startups that specialize in antibody design. Chai Discovery, a U.S. startup that received investment from OpenAI, attracted a 130 million dollar Series B investment in

December 2025, being recognized with a corporate value of 1.3 billion dollars, and is later known to be conducting negotiations to raise an additional 400 million dollars by increasing its corporate value to the 3.4 billion dollar level. The Chai-3 model released by this company signed a technology agreement with Pfizer, and it also announced a partnership with Eli Lilly in January 2026.

Looking at the industry as a whole, for one year from June 2025 to May 2026, over 980 million dollars in investment funds flowed through more than twenty deals solely into companies that design biological agents, including antibodies, using artificial intelligence. The investment fever is not cooling down even in the first quarter of 2026, with Arendel attracting a new 787 million dollars and EmendoBio attracting 80 million dollars.

Artificial intelligence models that draw the three-dimensional structures of antibodies are also noticeably boosting design speeds. Going one step further from existing technologies that predicted the three-dimensional structure when only the amino acid sequence of a protein was entered, reverse design has become possible, which inversely draws the three-dimensional structure of a binding site that perfectly fits the desired target shape and fills in the matching amino acid sequence. Rather than picking and wearing one of the clothes that already exist for the body, it is a method similar to a custom tailor shop that first takes the customer's measurements and cuts the fabric to match.

Such structure-based design techniques are being used not only for antibodies but also to create the dual-hook structures of nanobodies, peptides, and PROTACs, becoming a common foundational technology that penetrates multiple next-generation modalities.

Nanobodies, which are much smaller than antibodies, are also a field greatly benefiting from artificial intelligence. These mini-antibodies, found only in animals like camels and alpacas, are small enough to easily penetrate deep into solid tumor tissues, and they are also easy to mass-produce using

microorganisms. Recently, one research team formed a virtual research team made up of several artificial intelligence agents instead of people, and tasked them with designing from scratch a nanobody sequence inside a computer that neutralizes the novel coronavirus.

After the artificial intelligence agents went through the process of analyzing the structure of the target protein, suggesting binding candidates, and reviewing each other's designs, the researchers succeeded in actually synthesizing and verifying this sequence in a real laboratory. It is considered an example showing that if a human researcher sets a target and provides the direction, the artificial intelligence team can handle the rest of the design work almost autonomously.

A similar trend is visible in the field of designing peptides, which are small molecules made of a short chain of dozens of amino acids. An artificial intelligence tool called PepMLM, which borrows a natural language processing technique called masked language modeling, can design linear peptides that specifically bind to a target using only amino acid sequence information, without any three-dimensional structural information of the target protein. In the past, peptide design was a task that could only begin by precisely knowing the folded structure of a protein. However, as it has been confirmed that candidates with high binding strength can be selected without structural information, the scope of designable targets itself is expanding.

In situations where creating large molecules like antibodies or antibody-drug conjugates is burdensome, peptides are attracting attention as a lighter and cheaper alternative.

For an antibody or a CAR-T cell to recognize its target, there must be a sequence section acting like a handle to firmly attach to, which is called an epitope. This is similar to how one must find the exact door handle on a wide wall to open a door. The problem is that our body's immune system often mistakes the newly introduced therapeutic antibodies or cells themselves for

unfamiliar invaders and attacks them.

This is called immunogenicity, and if this side effect is severe, the drug loses its effectiveness or leads to a dangerous allergic reaction. Artificial intelligence scans the antibody sequences in advance to a degree that is hard for the human eye to measure, finds the sequence sections that would stimulate the immune system, and guides researchers to replace those sections with sequences already familiar to the human body.

Artificial intelligence tools that actually perform such immunogenicity predictions are also rapidly becoming more sophisticated. A prediction system called SITA, which calculates the immunogenicity score of the entire antibody sequence as well as the individual amino acid regions, clearly distinguishes antibodies derived from humans from those that are not. A multimodal deep learning model called ImmunoStruct, trained on binding data from about 26,000 peptides and major histocompatibility complexes, has raised prediction accuracy by combining sequence information, structural information, and biochemical properties all at once.

Tools like NetMHCIIpan, Hu-mAb, and BioPhi are actually being used in the humanization process to refine antibodies into sequences that are more familiar to humans. Because dangerous signals can now be filtered out using only the sequences long before a new drug candidate enters the animal testing stage, these tools have the effect of reducing both development costs and time.

Epitope prediction is also a core gateway when designing cancer vaccines. Even for the same type of cancer, the shapes of the mutant proteins appearing on the surface of cancer cells, known as neoantigens, are different for every patient. Researchers are using deep learning models that combine protein language models and convolutional neural networks to predict which neoantigens in an individual patient's tumor tissue will be well recognized by immune cells. These models are used to filter out candidates that will actually

cause an immune response from thousands of mutant proteins extracted from the tumor samples of melanoma patients. Based on these results, research to design customized cancer vaccines or immunotherapies just for one specific patient is underway at various universities and companies.

The prediction accuracy for CD8-positive T cell target epitopes presented by major histocompatibility complex class 1, which carries antigens to the cell surface, has also noticeably improved in recent years.

These artificial intelligence pipelines that find target epitopes and filter out immunogenicity in advance are not limited to antibodies. When designing the outer recognition part of a new artificial receptor attached to a CAR-T cell, the same type of prediction model is run to calculate in advance the risk that the patient's immune system might mistake the receptor itself for a foreign substance and cause a rejection response. Even for molecules closer to small molecules, like PROTACs, the possibility cannot be ruled out that new byproducts formed as they are metabolized in the body might trigger an unexpected immune response. Therefore, running metabolite prediction models alongside toxicity prediction models to filter out these risks in advance is now becoming a basic checklist item in the early stages of development.

From the stage of finding a target to the stage of verifying immunogenicity, the framework of the verification procedures performed by artificial intelligence is quite similar, even across different modalities.

Technologies that take a completely different approach in handling proteins that serve as drug targets are also growing rapidly. PROTAC is a name taken from the English acronym for proteolysis targeting chimera. Most existing drugs merely suppressed the movements of bad proteins. They were effective only while the drug molecules were attached to the protein, and the proteins would become active again as soon as the drug left the body or the protein structure changed even slightly.

PROTAC has completely abandoned this method. Instead of suppressing the activity of the bad protein, it hands the protein entirely over to the proteasome, which can be called the recycling disposal unit inside the cell, to be degraded. It is a double-hook molecule that grabs the protein to be eliminated with one hand, and with the other hand brings in an enzyme that attaches a degradation tag. The problem was that this double-hook structure is bulky and heavy, so it does not pass through cell membranes easily and its absorption rate in the body is also low.

Recently, machine learning-based pharmacokinetics prediction models such as EGFR-PROPK that target epidermal growth factor receptors have emerged, making it possible to fine-tune the molecular structure of protac molecules by pre-calculating how well they will be absorbed and distributed in the body.

One reason protac is receiving attention is that it can address targets that conventional drugs could not reach. Some proteins have smooth surfaces, so there is nowhere suitable for drugs to bind firmly. Drug developers have long referred to such proteins as undruggable targets. Because protac does not need to bind firmly enough to block activity - it only needs to catch onto the protein slightly, and then leaves the rest to the cell's own degradation machinery - many targets that drug developers had given up on are now returning to the candidate list.

Artificial intelligence is also being used to find binding sites where the protein can be barely caught. It scans the many curves and contours on the protein surface and proposes which areas should have which-shaped hooks attached, and also calculates how long a linker should be connected to that area so that the two proteins naturally come together.

Thanks to these artificial intelligence design tools, protac is showing results in actual clinical settings. Arvinas, which has been developing the breast cancer therapeutic bependogeston with Pfizer, presented the results of the Phase 3 VeriTAK-2 trial at the American Society of Clinical Oncology in 2025, and in

patients with previously treated, hormone receptor-positive, HER2-negative breast cancer, it statistically significantly extended progression-free survival compared to the existing therapeutic fulvestrant. The two companies applied for new drug approval from the FDA in June 2025 and received priority review designation. Chimera Therapeutics' oral STAT6 degrader KT-621 showed target protein degradation rates of 96 percent and 94 percent in blood and skin tissue, respectively, in a Phase 1b clinical trial for atopic dermatitis.

BeiGene's BTK-targeted protac BGB-16673 recorded a 78 percent response rate in an early trial in patients with previously treated, relapsed chronic lymphocytic leukemia, and then entered Phase 3 in 2025, and Bristol Myers Squibb's androgen receptor-targeted protac also began its first Phase 3 trial in prostate cancer patients. As of mid-2025, protac programs that have entered the clinical stage have reached twenty-eight globally, and among them, ten have already advanced to Phase 2 or Phase 3.

CAR-T cell therapy, which genetically engineers the cells themselves and uses them as a therapeutic, represents an even further advanced approach. The formal name means T cell therapy with a chimeric antigen receptor attached. T cells responsible for immunity are extracted from the patient's blood, genes are manipulated to insert an artificial receptor that recognizes only cancer cell surface proteins. It is similar to handing armed security guards a wanted poster containing specific facial features and sending them back out on patrol.

When these modified T cells are reintroduced into the body, they accurately find and attack blood cancer cells. While several CAR-T therapies have already been approved and are in use for blood cancers like leukemia and lymphoma, solid tumors that arise in the stomach, lungs, or colon are different. Solid tumors are surrounded by a thick barrier that prevents immune cells from accessing and functioning around the tumor, so even when CAR-T cells reach the area, they often cannot function.

Artificial intelligence is deeply involved in the design of CAR-T cells to overcome this solid tumor barrier. The phenomenon where receptors themselves incorrectly activate signals and become exhausted even before CAR-T cells enter the body is called tonic signaling, and artificial intelligence simulation models help by pre-calculating which receptor structures cause this early exhaustion and selecting more durable structures. Efforts are also continuing to automate the manufacturing process itself for creating CAR-T cells by combining digital twins that reflect each patient's individual physical condition with robotic equipment.

Before infusing cells into the patient's body, the design simulates in the computer how tumor cells and immune cells will actually fight, and how severely cytokine release syndrome - which causes signals to pour out all at once and leads to high fever and low blood pressure - will appear, and adjusts the cell design in a direction that lowers the risk.

In actual clinical settings as well, CAR-T cell therapies targeting solid tumors are showing results one by one. Allogene Therapeutics' off-the-shelf CAR-T cells ALLO-316 showed durable responses in an early trial in renal cancer patients and is being accepted as a signal that CAR-T cell therapies can work in solid tumors as well. In the STAR-101 Phase 1 trial, CAR-T cells targeting a new target called KIR showed safety and efficacy signals even at low doses in forty-two patients, and the results were particularly significant because they came from a cancer type where no cell therapies had previously been approved.

China's UTC Therapeutics received Chinese clinical trial approval for CAR-T cells using circular mRNA and presented Phase 1 results in solid tumors at the American Association for Cancer Research in 2026. In the United States and United Kingdom, new next-generation CAR-T clinical trials simultaneously targeting two targets have also opened in sixty pediatric solid tumor patients with conditions like rhabdomyosarcoma and Ewing sarcoma.

Despite these impressive design achievements, there are clear limitations. There is no guarantee that linker stability or binding affinity perfectly calculated in the computer will be reproduced exactly the same in actual living cells and animal models. There are still cases where unforeseen off-target effects or immune responses emerge later and lead to large-scale clinical trials being halted, and there are also cases where the data itself used to train AI models is biased toward certain races or specific clinical environments, making it difficult to apply directly to patients in different environments.

No matter how sophisticated the calculations in the design phase, the fact that the procedure of directly administering and confirming in the human body cannot be skipped weighs heavily over this entire field.

When looking at antibodies, antibody-drug conjugates, PROTACs, and CAR-T cells, a common flow can be seen in the role that artificial intelligence plays. Four tasks work together in a single cyclical structure: finding the proteins and epitopes that will be targets, designing molecules or cells from scratch that will attach to those targets, filtering in advance to see if they will stimulate the immune system, and checking in advance through simulation how they will actually move inside the body.

Unlike the drug development of the past, which intricately refined one type of small-molecule compound, the center of gravity is now shifting to the work of designing complex treatments mixed with multiple parts, such as antibodies, linkers, payloads, and genetically engineered cells, all at once from a multi-target perspective. Companies that possess both the artificial intelligence design capabilities and data to actually handle this work seem highly likely to take the lead in next-generation drug development.

Before a single antibody-drug conjugate or CAR-T cell therapy candidate that has reached the clinical trial stage today comes out, tens of thousands of candidate combinations have already gone through the process of being created and discarded inside computers. Before a needle actually enters a

patient's blood vessel, the procedure of first verifying the binding force of the antibody, the stability of the linker, the toxicity of the payload, and even the signal transduction of cell receptors in an in silico environment is solidifying as the standard, not the exception. A completely different industry is now being created compared to the drug development of the days when a single small-molecule pill was optimized.

The path that antibody-drug conjugates have walked has not always been smooth. Mylotarg, which was approved as a leukemia treatment in 2000, gained the reputation of being the first antibody-drug conjugate, but the linker attaching the drug was unstable, causing toxic substances to leak even into normal cells. Adding to side effects like veno-occlusive disease, it failed to prove a survival benefit again in subsequent clinical trials and voluntarily withdrew from the market in 2010. Even with Enhertu, which is widely used today, interstitial lung disease appears in about one out of ten patients. The bystander effect, where toxic substances leak beyond the target cells into surrounding tissues, is cited as the cause. In December 2025, worldwide clinical trials came to a halt all at once when ifinatamab deruxtecan, which Daiichi Sankyo and MSD were jointly developing, resulted in multiple deaths due to interstitial lung disease in a phase 3 trial for small cell lung cancer.

The US Food and Drug Administration lifted the hold, but the two companies had to introduce safety measures removing high-risk patients from the test subjects and retraining clinical trial personnel. European regulatory authorities have not yet opened the door. It seems that artificial intelligence cannot completely filter in advance how a linker, no matter how stably calculated on a computer screen, will react inside human lung tissues.

No matter how good the efficacy is, if the price is hard to afford, it ultimately will not reach the patient's hands. Antibody-drug conjugates like Trodelvy and Padcev cost millions of won per dose, and the patient bears that cost entirely until national health insurance coverage is confirmed. The Health Insurance

Review and Assessment Service's Drug Reimbursement Evaluation Committee acknowledged coverage while reviewing Trodelvy, based on the fact that there are no alternative treatments and it clearly increases survival time, even though the cost-effectiveness indicator came out worse than other drugs that had already finished review.

At a women's cancer policy debate held at the National Assembly Members' Office Building on May 19, 2026, Choi Seung-ran, the president of the Korean Federation of Breast Cancer Patients, said that metastatic breast cancer patients with genetic mutations stand before a treatment cliff because of costs and systems, even with treatments right in front of them. Even if artificial intelligence significantly reduces development time and costs in the design stage, unless those savings lead straight to drug price reductions, it is highly likely that the benefits patients actually enjoy will increase slowly.

The Korean government and companies are also keeping pace with this flow. On November 28, 2025, the Ministry of Food and Drug Safety released a guide for the first time summarizing what to look for when reviewing the quality of antibody-drug conjugates. It contained how to divide and write the quality data for the antibody, linker, and payload respectively, and what to review again when the manufacturing method changes.

Besides Ligachem Biosciences and Celltrion, IntoCell, which put forward its linker technology, is proceeding with joint development with Samsung Bioepis while signing a contract with China's Frontline Biopharma to develop two candidates together. Pino Bio handed over three of its five antibody-drug conjugate candidates to Celltrion and is continuing to develop the remaining two directly.

3.2

AI Decoding of Natural Product Drug Databases and Hybrid Discovery

In late autumn of 1971, in a laboratory on the outskirts of Beijing, Tu Youyou was turning the pages of *Zhouhou Beijifang*, a medical book written by Ge Hong around 340 AD. This experiment was not an individual research project, but rather part of China's national project that began in 1967, known as Project 523. As soldiers continuously collapsed from malaria, it was an enterprise in which the government gathered researchers from across the country and ordered them to find an antimalarial substance. At that time, malaria was a disease that took the lives of millions of people every year, and even strains with resistance to the existing treatment, chloroquine, were spreading widely.

Tu Youyou's team had already repeated the experiment hundreds of times but had failed to obtain any clear results. An old passage caught her eye, suggesting that sweet wormwood should not be boiled in hot water, but rather steeped at a low temperature as if squeezing out juice. She thus read from the ancient literature the fact that applying heat destroys the active medicinal ingredients. The substance born from this insight is the malaria treatment artemisinin, and Tu Youyou received the Nobel Prize in Physiology or Medicine in 2015.

People of ancient Egypt and Greece also knew from experience that chewing willow bark would soothe fever and pain. At the end of the 19th century, Felix Hoffmann, a chemist at the German company Bayer, chemically refined the ingredients in the bark and created aspirin. However, it was not until the 1970s, nearly 70 years after the drug went on the market, that the so-called mechanism of action - exactly which enzymes in the body aspirin blocks to soothe pain and fever - was scientifically identified. In a hospital laboratory in

London in 1928, Alexander Fleming returned from vacation and saw that a culture dish had been contaminated with blue mold.

Penicillin emerged from that accidental trace where bacteria failed to grow only around the mold. In the 1960s, botanists commissioned by the United States National Cancer Institute peeled the bark of the Pacific yew tree and found an anticancer ingredient, and this substance came out into the world as an anticancer drug named paclitaxel. It is known that a significant portion of the small molecule drugs sold in pharmacies today have their roots in structures derived from nature like this.

This list does not end at four. Digoxin, which comes from the leaves of a plant native to Europe called foxglove, has long been used as a drug to regulate the heartbeats of patients with heart failure. Vincristine, which is used to treat leukemia, came from an apocynaceous plant in Madagascar, and morphine, which is still considered a powerful painkiller today, came from the latex of poppy fruits. Statin class drugs that lower cholesterol also originally started from substances created by molds to protect themselves.

With just these few examples, we can guess the depth of time during which nature has served as humanity's medicine cabinet. Aspirin, penicillin, paclitaxel, and artemisinin were all the results of people observing nature for a long time and accumulating chance and perseverance.

Now, artificial intelligence is repeating this process of discovery on computer screens. It is estimated that there are at least hundreds of thousands of species of chemical substances created by plants, microorganisms, and marine life in nature, and if humans were to put each one on a laboratory bench to verify its medicinal effects, it would not be finished even if it took several generations. The space is so unimaginably vast for a person to search by hand that the metaphor that the number of theoretically combinable drug candidate molecules is greater than the number of stars in the universe is commonly used in the field of cheminformatics.

So far, the natural products that humanity has chemically verified and named amount to only hundreds of thousands of species, but if compounds whose structures have not yet even been revealed are added, it is estimated that the number will increase several times over. The technology that emerged to handle this vast space is the Graph Neural Network (GNN).

The artificial intelligence technology called a Graph Neural Network transforms a single molecule into a net structure, or a graph, where atoms are represented as dots and chemical bonds as lines connecting those dots, and it learns the pattern of this net as a whole. Just as a door opens when the uneven grooves of a key match the internal structure of a lock, a medicinal effect appears when the three-dimensional structure of a molecule matches a specific groove of a protein in the body, and a Graph Neural Network calculates this fit at the atomic level. It has been reported in recent research that a virtual screening model created in this way achieved a score exceeding 0.90 in accuracy indicators for sorting out compounds that actually have medicinal effects.

Models like DTIGN, which combine self-attention techniques, are known to have shown performance a step ahead of previous generation models in the task of predicting the binding affinity between compounds and target proteins. However, it is also pointed out as a limitation that because such models usually learn only with known binding information between compounds and targets, their predictive power can drop sharply when they encounter a completely new type of target that they have never seen during the training process. What used to take humans years to filter out hundreds of candidates in a laboratory, these models finish within a few days inside a computer.

As artificial intelligence reads the structures of phytochemicals - that is, chemical substances created by plants to protect themselves - one by one, previously unknown pharmacological mechanisms are also being revealed.

Exactly which signal transduction pathways in the body are regulated by berberine, a natural compound that has long been used for ulcerative colitis, is finally being drawn in detail through network pharmacology and artificial intelligence analysis. Fisetin, a quercetin derivative, is also having its target binding affinity and toxicity evaluated together through a quantitative structure-activity relationship model, that is, a model that calculates the mathematical relationship between a molecule's structure and its medicinal effect. Variational autoencoder models, trained specifically to handle the size, branching, and stereoisomeric structures of nature-derived molecules, have advanced to the point of generating new natural product-like structures that maintain complex three-dimensional frameworks while improving their properties as drugs.

The GNPS2 mass spectrometry database, built together by laboratories around the world, holds over 2.4 million spectra as of July 2025, and comparing the spectra of newly obtained samples against this database can prevent wasting time rediscovering already known compounds. What used to take people years to isolate a single active ingredient from hundreds of mixtures, artificial intelligence narrows down into a candidate group within a few hours just by inputting the structural formula.

All these predictions ultimately require well-organized databases to be supported. The public natural products database known as COCONUT collected natural product structure information scattered across literature worldwide in one place and organized it so anyone could download it, and the Natural Products Atlas compiled the source and production strain information of natural products created by microorganisms in a standardized format. If data remains scattered, even the most excellent artificial intelligence models cannot obtain materials to train on, so the work of continuously updating these public databases and filtering out errors, while hidden behind the announcement of brilliant new drugs, corresponds to essential foundational work without which nothing can proceed.

When mass spectrometry databases like GNPS2 are connected with these structural databases, we can narrow down what compounds are contained in an unfamiliar sample using only its spectrum.

This same decoding work extends beyond visible plants to the genomes of microorganisms in soil. Microorganisms store the blueprints of compounds they produce inside gene clusters called biosynthetic gene clusters, and the work of finding these clusters throughout the entire genome is called genome mining. A 2025 study on automated genome mining proposed a method to rapidly discover bioactive natural products using resistance genes from *Streptomyces* strains that protect themselves as clues. The newly developed genome mining tool NPtragM reportedly raised enzyme prediction accuracy to three times that of the widely used antiSMASH tool while reducing execution time to one-eighth.

Deep learning models such as e-DeepBGC and BiGCARP classify what types of biosynthetic pathways are hidden in a gene sequence and whether they are antibiotics or anticancer substances based on the sequence alone. It is known that most microorganisms living in environments difficult for humans to access, such as deep-sea hydrothermal vents or polar ice, cannot even be cultured in laboratories, but if we obtain only the genome sequence, artificial intelligence can read the blueprint of dormant compounds within it without culturing. It is known that among the strains for which genome sequences have been obtained, only a portion have revealed the actual products of the biosynthetic gene clusters they contain, so there is still wide room for genome mining to discover compounds in the future.

Since warnings have long been issued that without new antibiotics, an era will come when we cannot even treat common bacterial infections, the ability to read this chemical factory within soil is itself an urgent task.

Halicin and abaucin are both antibiotic candidate substances first discovered by artificial intelligence. The James Collins Laboratory at MIT is a place that

has responded to this urgency from an early stage. In 2020, this research team discovered through machine learning that halicin, a substance originally developed as a diabetes treatment candidate, unexpectedly possessed antibacterial effects that killed multiple bacteria. In 2023, they discovered abaucin, a substance effective against *Acinetobacter baumannii*, a bacterium resistant to antibiotics, and even clarified the mechanism of action: the substance inhibits the LolE protein, which is responsible for bacterial lipoprotein transport, thereby killing the bacteria.

In 2025, this research team departed from searching through existing compound lists and used generative artificial intelligence to draw new molecules that never existed before. After creating over 36 million virtual compounds on a computer and screening their antimicrobial properties, they discovered candidate substances targeting *Neisseria gonorrhoeae*, which is difficult to treat, and methicillin-resistant *Staphylococcus aureus* (MRSA), and these candidates had both different structures and different mechanisms of action from existing antibiotics. Based on this achievement, they received funding from the U.S. Advanced Research Projects Agency for Healthcare to design fifteen new antibiotic candidates, and they continue to pursue an antibiotic artificial intelligence project together with the nonprofit Fair Bio.

In October 2025, the same research team mapped out the exact pathway by which the newly discovered antibiotic candidate targets intestinal bacteria using generative artificial intelligence, and this was introduced as a case study that reduced the mechanism of action elucidation - which normally takes several years - to just several months.

Outside MIT, research on discovering antibiotics using artificial intelligence is proceeding in various ways. In February 2026, a scientific journal highlighted the research team of César de la Fuente, introducing how this team does not limit itself to compound lists or a single microbial genome, but uses artificial intelligence to broadly search through genetic information across diverse

organisms to discover short peptides with antimicrobial properties. The microbial communities living in human intestines and skin are themselves vast repositories of chemical information, and artificial intelligence is becoming increasingly skilled at finding short amino acid sequences that kill only pathogens within this information.

This perspective of not limiting the source of antibiotics to a specific plant or microorganism, but instead viewing genetic information from all organisms on Earth as a reservoir of candidates, is becoming reality only because we have artificial intelligence that can handle data.

Among nature's chemical diversity, one thick stratum is the record of traditional Korean medicine prescriptions accumulated over thousands of years. This prescription system, which mixes multiple medicinal materials in fixed proportions rather than a single herb and acts on multiple targets simultaneously, has long been treated only as empirical medicine and has had difficulty receiving scientific recognition. When reverse engineering techniques - which deduce design principles by working backward from the finished product - are applied here, we can trace back one by one to see which targets each medicinal material in the prescription addresses and how their combined strength produces synergy. To experimentally verify a single component that slows aging would require testing hundreds of prescriptions and thousands of compounds one by one on cells, but artificial intelligence filters through this process first using data.

The artificial intelligence drug discovery company Insilico Medicine created a foundation model called Precious3GPT by training it on over 1.2 million pieces of omics data, which are experimental data measuring genes, proteins, and metabolites in large quantities extracted from the tissues of various species, including humans, mice, and macaque monkeys. Such a model is called a foundation model to mean that it was trained broadly in advance so it can be used across multiple tasks, and Insilico Medicine combined this model with the

BATMAN-TCM2 database, which collects traditional medicine ingredients and target information. This combined model found 13 aging-related genes that appear commonly across humans and other animal species, and it selected 34 traditional medicine ingredients that overlap with these genes.

Every time a cell divides, the telomeres at the ends of the chromosomes gradually shorten until the cell itself eventually ages, and as a result, it was confirmed at the molecular level that various ingredients contained in an old prescription called Hua Shan Wu Zi Dan work together to slow this telomere shortening in liver, lung, and muscle tissues.

Attempts are also continuing to go beyond reading a single prescription and to turn the entire knowledge of traditional medicine into a searchable map. Through literature mining, the database TCMBank has organized the connections among 9,192 types of traditional medicinal materials, 61,966 ingredients, 15,179 target proteins, and 32,529 related diseases. Not stopping there, it connected the three-dimensional structure of each ingredient with Western new drug databases like DrugBank and PubChem, making it possible to compare at a glance even the chemical similarity between traditional medicine ingredients and already approved synthetic drugs. Thanks to these comparisons, researchers can proceed with drug repositioning work, which re-examines a drug already used for another disease as a candidate for a new indication simply because its structure resembles a traditional medicine ingredient, without having to synthesize a new compound from the beginning.

TCM-GPT is the first 7 billion-parameter language model pre-trained solely on traditional medicine literature, and it achieved scores that exceeded existing general-purpose language models on traditional medicine doctor licensing exam questions and diagnostic tasks. Shuimu Molecular, an artificial intelligence company in Beijing, joined hands with the AIR Institute at Tsinghua University to release an interactive drug discovery assistant called ChatDD based on a large language model with 100 billion parameters. It

provides a module to help discover targets and search for candidate substances separately from a module to help design clinical trials, and it is expanding its application to the modernization research of traditional medicine. Problems that could not be solved by Western medicine's single-target approach, which looked at ingredients, targets, and diseases separately, are opening up again as artificial intelligence translates into numbers the principles of traditional medicine, where multiple ingredients are tangled like a net to press multiple targets at the same time.

A research team at the University of Macau has built a separate artificial intelligence platform that integrates traditional medicine literature and compound data, targeting neurodegenerative diseases like Alzheimer's disease where multiple causal targets are entangled, and they are testing whether a multi-target approach can work for these intractable diseases as well.

Despite these achievements, it is not easy for artificial intelligence to swallow traditional medicine data just as it is. Diagnostic information accumulated by taking the pulse, looking at the tongue, and asking about symptoms was not recorded in numbers in the first place, and prescription records are also written differently at each hospital and clinic, so it takes considerable effort just to organize them into a single table. This is also why researchers who build databases read through papers and old books one by one and verify the ingredient and target information twice by human hands.

If non-standardized data is forced into training, artificial intelligence carries the risk of presenting nonexistent correlations as if they were real, so the task of refining the data is just as important as developing a flashy model.

The Korea Institute of Science and Technology digitized its own natural product sample information and developed an artificial intelligence platform called NPI-finder that predicts protein targets. It proved the efficacy of a persimmon leaf ethanol extract in improving dry eye symptoms and even led

to approval from the Ministry of Food and Drug Safety and a technology transfer. MThera Pharma, founded by CEO Son Mi-won in 2020, built SyMthomics, a big data platform that precisely uncovers the mechanisms of action of natural products made up of multiple ingredients. A candidate substance for treating diabetic neuropathy that came out of this platform achieved a result of reducing pain by 48 percent in a phase 2 clinical trial in the United States and is preparing for phase 3, while a candidate substance for treating Parkinson's disease finished a phase 1 clinical trial in the United States in 2022 and received interest from global pharmaceutical companies like AbbVie and MSD.

A research team led by Professor Choi In-ho of the Department of Medical Biotechnology at Yeungnam University announced in 2026 that they had found a candidate natural product substance to treat sarcopenia by combining artificial intelligence with in silico analysis. They exactly followed the sequence of narrowing down candidates through computer analysis alone without experiments, and then confirming them through cell and animal experiments. The government is also adding strength to this trend by establishing the artificial intelligence and digital great transformation of traditional medicine as a core goal in the 5th Comprehensive Plan for the Promotion and Development of Traditional Medicine, which will run from 2026 to 2030. Natural product new drugs, which for a while were evaluated as being pushed to the background in the domestic pharmaceutical industry, are emerging as an asset receiving the spotlight once again thanks to artificial intelligence, which greatly reduces discovery costs and time.

No matter how many candidates artificial intelligence picks out, the final confirmation must still go through petri dishes and animal models. Even after a research team at the Massachusetts Institute of Technology narrowed down candidates expected to have antibacterial effects from 36 million virtual compounds, experiments to actually synthesize the compounds, expose them to bacteria, and watch if they survive were still necessary. The Korea Institute

of Science and Technology's NPI-finder also did not end with just computer predictions, and it was only able to receive approval from the Ministry of Food and Drug Safety after going through a testing stage of actually applying the persimmon leaf extract.

This cycle of alternating between prediction and experiment, where the error is fed back into the model whenever the experimental result deviates from the prediction, is called hybrid discovery. The division of roles, where artificial intelligence first narrows down a search range that a human could not handle, and humans manually check only that narrowed list, is now the standard in this field. Only when we accept that each prediction made by artificial intelligence is not an immediate answer, but a hypothesis to be confirmed by experiment, does this cycle truly move forward to the next candidate substance without spinning its wheels.

When phytochemical data and traditional Korean medicine prescription data meet deep learning and binding affinity prediction algorithms - algorithms that calculate how strongly a drug molecule sticks to a target protein - a new design method called nature-mimicking structure is born. Instead of drawing a chemical compound from a blank slate, this is a method where artificial intelligence learns the frameworks that nature has refined over hundreds of millions of years, modifies those frameworks, and proposes new candidate substances. While it took decades of accumulated chemical knowledge for Felix Hoffmann to refine a single willow bark into aspirin, we can now simultaneously draw thousands of variants of a framework in the same family and select the most promising candidates within a few days.

It has been reported that a model created by refining the recently released AlphaFold3 series of structure prediction models for the task of scoring the binding strength between chemical compounds and proteins achieved a similar level of accuracy to free energy perturbation calculations, which previously took days, with much less computation. Considering that many

already approved treatments, including antibiotics, anticancer drugs, and lipid-lowering drugs, are ultimately substances that branched out directly or indirectly from frameworks originating in nature, the current trend where artificial intelligence learns the grammar of those frameworks to propose new structures seems like a natural next step.

The reason natural products are special as materials for new drugs is that they are structures already filtered once by hundreds of millions of years of evolution. Plants and microorganisms have created chemical substances to deter predators or suppress competitors, and in that process, three-dimensional structures refined to interlock with the proteins of living cells have naturally accumulated. Unlike chemical compounds designed from scratch on a computer in a laboratory, which often fail to pass through cell membranes or break down quickly inside the body, the frameworks of natural products are structures that have already passed the test of surviving in a biological environment.

The fact that several drug classes currently used in hospitals, such as antibiotics, anticancer drugs, and immunosuppressants, branched out from these naturally derived frameworks supports this strength. Among chemoinformatics researchers, the view that new drug candidates have a higher probability of surviving to the clinical stage when starting from a framework found in nature, rather than from a randomly generated virtual compound, is gaining ground.

However, for the same reason, natural products are difficult to handle. Hundreds of different components are mixed within a single plant, making it hard to identify which component produces the medicinal effect, and it is common for the active ingredient to exist at a level of only one millionth of the total extract. The component content changes from year to year depending on the climate and soil of the native habitat, making it difficult to secure reproducibility, and molecules with complex structures are not easy to

chemically synthesize in large quantities. In fact, when paclitaxel was first produced, the bark of thousands of Pacific yew trees was stripped, bringing the tree itself to the brink of resource depletion. The paradox of driving the source plant or bacterial strain to the brink of extinction while trying to harvest the active ingredient has still not been completely resolved today.

The classic exploration method has long been criticized for its limitation: it often repeatedly finds the same chemical compounds, meaning the probability of encountering a truly new structure is low no matter how many samples are collected. Artificial intelligence reduces this waste of duplicate discovery by pre-filtering the mass spectrometry fingerprints of already known chemical compounds, but obtaining the discovered substances again in sufficient quantities from actual plants and microorganisms still requires human hands.

The use of artificial intelligence is also growing in predicting how natural products are absorbed, broken down, and cause toxicity when they enter the body, the so-called ADMET prediction. Extracts where multiple components are mixed together involve complex interactions where components help or hinder each other's absorption, making it difficult to calculate toxicity as one would with a single compound. Recently, multi-component prediction models that calculate the probability of liver toxicity or heart toxicity by taking not only the structure of individual components but also their combination ratios as inputs are being announced one by one.

For natural product drug developers aiming to get regulatory approval for original complex extracts rather than purified single components, this predictive ability is becoming a practical tool that lowers the probability of failure right from the clinical trial design stage.

This work of converting knowledge obtained from nature into digital data also brings up the question of whose knowledge it originally was. The Nagoya Protocol, adopted in 2010, is an international agreement that mandates that if

profits are made by using a country's biological resources and the traditional knowledge attached to them, those profits must be shared with the resource's country of origin and local communities. The problem is that what artificial intelligence deals with these days is not physical resources like plant specimens or seeds, but digital sequence information that transcribes their genetic sequences. Seeds pass through customs, but data does not. As it became possible to train artificial intelligence simply by downloading sequence information from public databases without crossing borders with a single biological sample, a long debate has continued over how to share profits with the countries of origin.

The anticancer drug vincristine, which came from a plant in the Apocynaceae family in Madagascar, is often cited as an example illustrating this concern. While a Western pharmaceutical company developed an anticancer drug that generated massive sales from this plant, it has long been pointed out that the country of origin, Madagascar, received almost no fitting compensation. At the negotiating table, developing countries rich in genetic resources have argued that digital sequence information should be treated the same as physical resources, while pharmaceutical companies and academia in developed countries, which mainly handle this information, have countered that overly strict regulations would instead slow down the pace of research.

On February 25, 2025, in Rome, Italy, parties to the Nagoya Protocol under the Convention on Biological Diversity launched an international fund called the Cali Fund, and two days later also adopted an official decision on digital sequence information. Companies that gain profits from using digital sequence information of genetic resources are recommended to contribute 0.1 percent of their sales or 1 percent of their profits to this fund, and the collected funds are used to support developing countries and indigenous communities rich in biological diversity.

Although few companies currently follow this recommendation and it lacks strong legal binding force, it holds significance as the international community's first step toward returning profits in some form to countries of origin in an era when new drugs are designed using digital information alone. This fund reaffirms with numbers the fact that underlying all work to reverse-engineer herbal prescriptions and mine microbial genomes are biological resources long maintained by specific lands and specific communities.

Converting a single herbal prescription into a graph and inputting a single microbial genome into a language model is now technically not difficult. What actually takes time is the final stage of transforming candidate substances discovered this way into a medicine that can be safely administered to real patients, and the process of honestly sharing profits from the knowledge and biological resources that formed the foundation of that discovery. Both graph neural networks and large language models ultimately operate on data accumulated through thousands of years of human observation, documentation, and failure. Just as an old passage about brewing sweet wormwood at low temperature led to a Nobel Prize, any single prescription or sequence now dormant somewhere in a database could lead to the next new drug.

Now that artificial intelligence reads records left by hundreds of millions of years of evolution and thousands of years of human experience, whether fair profits return to the lands and peoples who first created those records appears likely to determine the success or failure of this technology going forward. The fact that a single medicine comes to completion only when everything is connected - the graph on the screen, the culture dish on the laboratory bench, and the forest and village where its roots grew - appears unlikely to change no matter how sophisticated artificial intelligence becomes.

Over the past 10 years, the Korea Institute of Oriental Medicine has gathered clinical data from nearly 50,000 patients through artificial intelligence-related projects and genetic epidemiology studies. Diagnostic methods relying on the experience and subjective judgment of Korean medicine practitioners have long been identified as an obstacle in efforts to have Korean medicine recognized as a science. The institute chose instead to gather genetic data, facial photographs, and wearable device measurements together to support such judgments with numbers.

Beyond this, the Korean medicine community has also begun on its own to organize clinical records according to FHIR format, an international medical information standard. Following this standard makes it possible for even small medical records from Korean medicine clinics to be linked with other medical information such as hospital electronic medical records or health insurance data. The institute plans to use the data gathered in this way to develop Korean medicine precision medical technology.

Standardizing components remains an obstacle to crossing the regulatory threshold. In 2015, the Board of Audit and Inspection pointed out that the natural product new drug system provided excessive benefits to domestic companies, and subsequently the Ministry of Food and Drug Safety eliminated the separate classification of natural product new drugs altogether and raised examination standards to the level of new drugs. The U.S. Food and Drug Administration also issued guidelines for natural product medicine development in 2016 in an attempt to address this issue, but cases of plant extracts receiving approval as drugs without processing remain few - limited to examples like Veregen in 2006 and crofelemer in 2012.

This is because it is not easy for an examiner to determine each time that a substance whose component content varies with the collection period and origin is the same drug. Even if artificial intelligence calculates component variation in advance and provides a correction value, it seems likely that

human examiners will continue to be responsible for determining approval based on that value.

One case in which traditional knowledge led to a new drug offers insight into how to think about sharing profits. Indigenous peoples of the Peruvian Amazon have long used the red sap from the *Croton lechleri* tree as medicine for wounds and stomach upset. A pharmaceutical company that recognized this knowledge has maintained for decades a relationship of sustainably harvesting sap from 19 indigenous communities at some 30 locations. Crofelemer created in this way received U.S. Food and Drug Administration approval in 2012 as a medicine for chronic diarrhea in AIDS patients, and the company reports that in return it has supported education and water supply projects in the harvesting areas.

In other words, trust built over decades has supported this kind of sharing more than any single contract could. However, in an age like now when new drugs can be designed using sequence information alone, it appears that relying solely on the goodwill of companies will not be sufficient to protect communities of origin over the long term.

3.3

Nanomedicine and Smart Targeted Drug Delivery Systems

In December 2020, cargo planes carrying special freezer containers capable of withstanding temperatures below 70 degrees Celsius landed one after another at airports in the United States and Europe. Inside the boxes was the COVID-19 vaccine BNT162b2, jointly developed by Pfizer and BioNTech, and within a single drop of that vaccine floated billions of invisible oil droplets. Each of these oil droplets is precisely a lipid nanoparticle (LNP). Moderna completed its initial design just 25 days after the gene sequence was made public. Pfizer and BioNTech received emergency authorization at the record-breaking speed of eight months. Yet what ultimately decided success for both was not the messenger ribonucleic acid (mRNA), the core material of the vaccine, but rather the oil droplet - the nanoparticle - that safely delivered that delicate genetic material into the body.

mRNA was a weak molecule that would quickly be broken down by enzymes in the body on its own. If this oil droplet had not wrapped around it like an umbrella, the vaccine would never have come into the world. From that day onward, billions of people around the world rolled up their sleeves and received this oil droplet injection. Lipid nanoparticles rose overnight from an obscure laboratory material to humanity's first commercially applied nanomedicine, experienced bodily by the entire human population. At the root of this oil droplet technology lie Pieter Cullis, who had researched lipid nanoparticles for decades in Canada, and Acuitas Therapeutics, the company he founded. A substantial portion of the core patents for the lipid nanoparticles used by Pfizer and BioNTech came from this Canadian company.

What landed on that runway that day was not merely a box of vaccines, but the moment when nanomedicine crossed over the laboratory fence and entered human arms worldwide.

One nanometer is the size of one meter divided into one billion parts. If we call one glass bead one nanometer, then one meter would be the size of one Earth - such is this tiny world. The lipid nanoparticles used in the vaccine have a diameter of roughly around 100 nanometers, less than one-thousandth the thickness of a human hair. A single red blood cell in our body is roughly around 7,000 nanometers across. If we line up these nanoparticles in a row, just over seventy are needed to cross a single red blood cell. All the stories of nanomedicine begin at a scale so small that individual particles cannot be seen with the naked eye or even with an ordinary microscope.

Yet in such an extraordinarily small world, if the size differs by only a few nanometers, it completely changes whether a particle goes to the liver or lungs, enters inside cells or is filtered out from outside. What molecules attach to the surface matters equally. Whether the surface carries a positive or negative charge is likewise a decisive factor that changes the destination. Just as postal packages are classified into different counters according to their size and surface markings, nanoparticles are sorted into different organs throughout the body according to their size and surface properties. The sizes of the gaps in blood vessels and cell membranes throughout the body also vary. If a particle is too large, it cannot enter at all. If it is too small, the kidneys filter it out and it quickly passes out of the body.

The idea itself of using nanoparticles to deliver medicine is not new. In 1995, the United States Food and Drug Administration approved a drug called Doxil, which wrapped the anticancer drug doxorubicin layer upon layer with an oil membrane. This marked the first case of nanotechnology actually being used in real patients. Subsequently, follow-up medications such as Abraxane, which attached the anticancer drug paclitaxel to albumin, a blood protein, did

emerge. But progress remained slow. As of 2025, all nanomedicines approved for clinical use worldwide combined amount to only around 50 to 80. Moreover, most did not deviate much from chemical principles that are decades old.

It took several months to create a new combination in a test tube, introduce it to cells and animals, and obtain results. If results proved unsatisfactory, one had to start completely over from the beginning. This slow pace, which for nearly 30 years had relied on human intuition and repeated experimentation, finally began to accelerate when it met artificial intelligence.

In 2025, an artificial intelligence agent named NanoSafari read through more than 20,000 nanoscience papers without stopping, day and night. This agent gathered nanoparticle synthesis conditions, sizes, and surface property data scattered throughout the papers to create a knowledge database connected to a large language model. What a single researcher must read over a lifetime, artificial intelligence accomplishes in one night. A literature review that once took months now takes a single day to complete. Previously, researchers had to spend months or years at the laboratory bench in trial and error to find the optimal combination for even a single nanoparticle.

Now an approach called nanoarchitectonics - constructing structures with desired functions by stacking atoms and molecules one by one like bricks - is spreading rapidly with the aid of artificial intelligence. Just as architects draw blueprints before constructing buildings, it is becoming standard for artificial intelligence to first draw molecular blueprints and then assemble nanoparticles according to those blueprints. This design method extends beyond new drug development to the creation of cosmetics and fine chemical materials, expanding as an approach that encompasses nano-scale design in general.

A paper published by researchers from Sungkyunkwan University in the international academic journal *Advanced Materials* in 2025 clearly illustrates

this trend. It forms a single ring: beginning with using bioinformatics to identify disease-causing targets in the body, proceeding through the stage of precisely designing nanoparticle surfaces using machine learning, and continuing to the stage of calculating in silico - that is, within a computer - where these particles will disperse throughout the body and when they will release the drug. This ring does not end in one direction but returns to the beginning, forming a closed circular structure that is reflected in the next design.

In the past, a single cycle of design, synthesis, and evaluation would take several weeks, but inside a computer, it can be repeated many times in just a day or two. As this cycle becomes tighter, off-target side effects - where drugs accumulate in the wrong organs - are reduced, and designs tailored to individual patients' physical conditions are becoming much closer to reality. By shifting the trial and error once repeated on laboratory benches to simulations on computer screens, this appears to be a turning point that fundamentally changes the pace of nanomedicine research.

If you look closely at a single nanoparticle designed in this way, it resembles a carefully wrapped gift box. Attached to the outermost shell are targeting ligands that stick only to specific tissues or cells; these are small tags, like proteins or sugar chains, that attach precisely to specific receptors in the body. These tags recognize demanding destinations, such as the tumor microenvironment (TME) or the blood-brain barrier (BBB), and open the door. Inside that lies a scaffold where the thickness, arrangement, and composition ratio have been meticulously calculated by machine learning, determining the overall size, rigidity, and degradation rate of the particle.

And in the innermost core is the drug that actually acts on the body, the payload, which only begins its work after arriving at the destination. The key point is that the three layers can be designed separately; by leaving the inner materials as they are and only changing the outer tags, a new nanoparticle targeting a completely different organ can be created. If the design of any one

of these three layers goes wrong, the nanoparticle will leak the drug in the wrong organ or be washed out of the body before even reaching its destination.

After a deep learning platform named Agile learns on its own from a virtual lipid compound library of around 60,000 entries, it refines the data with measurements from large-scale synthesis in actual laboratories to determine in advance how well ionizable lipids will transfer genetic material into cells. In August 2025, the international academic journal Nature Nanotechnology introduced a neural network model with a transformer structure named COMET, which learned from the largest-ever lipid nanoparticle dataset named LANCE and was able to predict the efficacy of unfamiliar lipid nanoparticle combinations.

In November of the same year, a research team at Johns Hopkins announced that after creating and testing hundreds of lipid nanoparticle formulations at high speed, they used machine learning to discover formulation conditions that improve gene editing efficiency. In 2026, a system named LipoAgent also appeared in the academic community, in which multiple large language model agents collaborated by dividing roles to propose new lipid structures and cross-checked each other's designs to consider safety as well. These are examples that clearly show the trend of laboratory benches turning into data factories and computers becoming designers.

A technology called selective organ targeting (SORT) is widely used to determine which organ to send lipid nanoparticles to. Just as different doors open depending on which key is chosen from a keychain, the organs the particles seek out change depending on the properties of the lipid molecules added to the surface of the lipid nanoparticles. Experiments have confirmed that adding positively charged lipids causes the particles to flock to the lungs, adding negatively charged lipids sends them to the spleen, and adding ionizable amino lipids drives them to the liver. The sequence is that as the

polyethylene glycol component covering the particle's surface falls off inside the body, specific proteins in the blood newly attach to that position, and as these proteins pair with the receptors of each organ's cells, the destination is determined.

As it has become possible to select and target organs, the side effects caused by gene therapies that should only operate in the liver spreading to the lungs or heart are also decreasing. Instead of humans testing these subtle chemical combinations one by one by hand, artificial intelligence is taking on the role of rapidly filtering them out from among numerous candidates.

Carriers of the polymer and micelle series are becoming as sophisticated as lipid nanoparticles through the hands of artificial intelligence. Machine learning models that have jointly learned from microfluidic engineering equipment that creates particles in minute fluid channels and high-resolution imaging data can adjust the size, surface charge, and even the shape of the particles on their own. The phenomenon where proteins in the blood plasma flock to and stick to the surface of a nanoparticle the moment it enters the bloodstream, creating a new surface as if wearing a coat, is called a protein corona, and depending on what this coat is woven from, how long the particle stays in the blood and which cells it is absorbed into are determined.

Based on a database collecting 817 nanoparticle formulations and 2,497 protein adsorption data points, researchers trained interpretable algorithms like Random Forest or XGBoost, which reach conclusions by gathering multiple small decision rules, and revealed that particle size, surface charge, and the time spent in the blood are the core variables that dictate the protein corona. Recently, research predicting the protein corona in much more complex nanostructures made by folding DNA strands is following, showing a trend where the targets of prediction are expanding beyond lipids and polymers.

As these predictions become more sophisticated, cancer drug carriers that need to stay in the blood for a long time and vaccine carriers that need to be

rapidly absorbed into target cells can be made separately with different surface designs. This means it has become possible to draw the shape of this coat in advance inside a computer before entering clinical trials.

When the brain becomes a target for treatment, a much higher barrier awaits. The walls of cerebral blood vessels are surrounded by a dense mesh-like blood-brain barrier (BBB), and among 100 oral therapeutic candidates, only at most six, and at least two or three, pass through this barrier. Among new drug candidates designed to target Alzheimer's disease or Parkinson's disease, there are countless cases where they show effectiveness in animal experiments but fail to cross this barrier in human bodies and collapse at the final stages of clinical trials. A predictive model that combines MegaMolBART, a transformer-based model, with XGBoost learned from a database organizing whether about 8,000 compounds pass through this barrier, and achieved a score of 0.88 on the area under the curve metric, which shows how well predictions match reality.

In August 2025, American Lantern Pharma released an artificial intelligence module called predictBBB.ai, which can identify whether a compound passes through the blood-brain barrier with accuracy reaching 94 percent based on just its structural information. These predictive models, which read molecular structures through graph neural networks, serve as gatekeepers that filter out new drug candidates on computer screens before they enter animal experiments. Thanks to the failed candidates filtered out by one such prediction tool, there is an evaluation that the budget and time that used to go into animal experiments are being significantly reduced.

Large molecules that are chemically difficult to pass through this barrier take a completely different path. By attaching antibodies or small proteins that bind to transporter proteins like transferrin receptors, which appear abundantly on the surface of cerebral blood vessel cells, to the surface of nanoparticles, they can, like soldiers hidden inside a Trojan horse, subtly ride along with the

substance transport work that the receptor originally performed and be carried deep into the brain. Artificial intelligence is being used to find, among numerous amino acid combinations, which parts of the antibody should be changed and how so that it binds tightly to the receptor while not causing unexpected immune reactions in the body.

This receptor-mediated transport technology, paired with predictive models, now serves as a new standard, with the sequence of first filtering on computers which antibody designs significantly increase the probability of crossing through, then confirming through experiments.

Facing this barrier that is difficult to penetrate with synthetic nanoparticles alone, what scientists paid attention to is exosomes, nano-sized pouches that living cells emit from within themselves. Because these are vesicles that the body originally creates and uses, the risk of the immune system mistaking them for foreign substances and attacking them is lower than with synthetic nanoparticles, which is one reason why exosomes are preferred in clinical practice. Clinical trials are being conducted at several sites where medications are filled into exosomes obtained from stem cells extracted from fat or mesenchymal stem cells from bone marrow at an optimal ratio calculated by artificial intelligence, and then administered by spraying through the nose. This is because when exosomes enter through the bundle of olfactory nerves that contacts the nasal mucous membrane, they can reach brain lesion areas directly without going through blood circulation that travels throughout the body.

In other words, it is a pathway that goes straight to the brain via a shortcut called the olfactory nerve without passing through the major highway called blood. Experiments have also been reported where protein delivery technology that turns operation on and off with light or CRISPR interference technology that silences gene expression are loaded onto exosomes, turning off the genetic switch of alpha-synuclein, a protein that causes disease in the

brain neural networks of experimental mice with Parkinson's disease. Because they borrow carrier vehicles that the body originally creates and uses, exosomes are considered the next-generation carriers that will deliver drugs to the brain.

However, there is still some distance before exosomes leave the laboratory and appear on pharmacy shelves. As of 2026, there are around 120 exosome therapeutic candidates in development worldwide, and about 40 percent of them have entered clinical trial phases, but unlike market projections estimated at a scale of 58.1 billion dollars, no exosome therapeutics have yet received approval from the U.S. Food and Drug Administration. A candidate substance called ExoPTEN targeting spinal cord injury or optic nerve damage is also going through preclinical phases with the goal of entering human clinical trials, and the progress of subsequent candidates is also active.

Capricor Therapeutics, which has been developing Deramiciel, a cardiac-derived cell therapy that secretes exosomes, received a 9-to-3 negative vote at a Food and Drug Administration advisory committee meeting held on July 29, 2026, stating that it failed to prove its effectiveness against Duchenne muscular dystrophy cardiomyopathy, and is awaiting the results of the final review scheduled for August 22. Although advisory committee votes do not have binding force, this case demonstrates that even as sophisticated artificial intelligence predictions and animal experiment results accumulate, cautious standards are still applied before the final gate for human trials.

The industry predicts that regardless of how the final conclusion on Deramiciel turns out, the result will have a long-lasting impact on how regulatory authorities view the entire therapeutic approach called exosomes.

Turning the eye to solid cancers rather than the brain, another challenge called the tumor microenvironment (TME) awaits. The tissues surrounding tumors have blood vessels extending in crooked paths, lack oxygen, and have acidity different from normal tissues, making it difficult to deliver drugs to every

corner with just one approach. The phenomenon where nanoparticles permeate well into tumor tissues with unusually sparse gaps in blood vessel walls and remain there for a long time is called the enhanced permeability and retention effect, and research teams put biological response information extracted from 378 tumor datasets and about 200 papers into both a structure-activity relationship model containing the relationship between nanoparticles' chemical structure and biological effects and a physiological pharmacokinetics model that mathematically expresses the process of how drugs are absorbed and distributed in the body, calculating this phenomenon based on 24 hours later.

The accuracy was so high that the coefficient of determination, which shows how well the predicted values and actual values match, reached 0.83. Going one step further, biomaterial reverse engineering completely flips the order around. If you first input a final goal, such as wanting to kill tumor cells or wake up immune cells, the algorithm goes back through a vast database and finds the polymer ratio, surface ligands, and even camouflage coatings mimicking tumor cell membranes or red blood cell membranes that fit that goal in reverse. This is how smart carriers are being made that react only to specific enzymes or the low acidity released by a tumor, bursting the drug only in that location.

However, because the detailed conditions of the microenvironment are different for each tumor, and even for the same tumor in each patient, it is often not easy to transfer a design proven in one patient exactly as it is to another patient.

Just as much as sending treatments into the body, nanotechnology and artificial intelligence are working together to find out early on how deep a disease has progressed. Nano-optical technologies like surface-enhanced Raman scattering or localized surface plasmon resonance can catch even the signals of a disease floating in very small amounts in the blood or

cerebrospinal fluid and amplify them in the form of light. Simply put, it is similar to the principle of gathering the faint light of a flashlight with a concave mirror and shining it brightly. Artificial intelligence reads this amplified, complex spectrum of light and fluorescent signals in real time to find out the early moments when amyloid beta and tau proteins, which are markers for Alzheimer's disease, begin to clump together, earlier than anyone else.

It is different in nature from existing diagnostic methods in that it finds the seeds of a disease years before symptoms appear on the surface, at a time when even the patients themselves do not notice. If high-resolution nano-imaging equipment and deep learning image analysis are used together, you can watch from outside the body how much the nanoparticles put into the body have actually gathered at the lesion area, and at what speed the drug is flowing out.

Taking a step further from this, autonomous precision medicine is emerging as the next goal, where nano sensors planted in the body send the patient's biosignals back to artificial intelligence in real time, and the artificial intelligence adjusts the drug release speed itself to match those signals. Just as devices that measure blood sugar and automatically adjust the amount of insulin are already staying by the side of diabetes patients, there are predictions that the day is not far off when nano sensors and nano carriers in the body will pair up in cancer or degenerative brain diseases to adjust the amount of drugs all day long without going through human hands.

However, for such an autonomous system to enter actual hospital rooms, safety devices must also be verified together to prevent too much medication from being released when a sensor sends the wrong signal. Nano systems that contain diagnosis, treatment, and the autonomous regulation function connecting the two in one body are chosen as examples showing the direction nanomedicine will go next.

High-quality nanotoxicity data enough to train artificial intelligence models is extremely lacking, and because the measurement methods and standards are different for each laboratory, it is not easy starting from the work of gathering the data together. It is not rare for toxicity numbers to change depending on which laboratory measured them even for the same nanoparticles, so there are also cases where it is difficult to judge whose side to take among different results. Even the work of gathering data that various universities and companies have built up with different protocols into a single standard is staying at the toddler stage right now. No matter how high the prediction accuracy is, deep learning models have a black box problem in that they cannot clearly explain to clinicians or regulatory officials why they judged that a specific nanoparticle structure does not cause toxicity.

If traditional biophysical models that solve the movement of materials with differential equations and artificial intelligence models that find answers directly from data do not match their hands and feet together, there is always the danger remaining that predictions that were perfect on a computer screen will spread into immune rejection reactions or unexpected side effects inside the actual body. Moreover, even if it is a nanoparticle structure newly proposed by artificial intelligence, you still have to bear the costs of preclinical and clinical trials over several years until it is put into a human body, so just because the design stage has become faster does not mean the entire timetable for new drugs to come onto the market will shrink by that much.

The clock of the system is not exactly catching up with the speed of technology. The fact that more than half of the 46 new drugs approved by the US Food and Drug Administration in 2025 were treatments with world-first mechanisms shows that regulatory authorities are also changing their review methods in their own ways, but in cases where the material itself is made up of multiple layers like nanomedicines, a separate standard is still needed. The US Food and Drug Administration passed the FDA Modernization Act 2.0 to revise the long-standing rule that animal testing must be passed, and opened the

way to replace preclinical toxicity testing with organ-on-a-chip or computer simulations.

However, regulatory agencies in various countries, including the European Medicines Agency, are asking for different review standards for advanced nanomedicines crossing the blood-brain barrier or carriers coming from living cells like exosomes, so it is common for a product that passed in one country to have to prepare data again from the beginning in another country.

It is also not yet clear where the line is on how far personal information must be consented to during the process of artificial intelligence learning a patient's genome information in large quantities. When nanoparticles designed and even dosed by artificial intelligence itself cause serious side effects to a patient, there is also no set answer on who among the software developer, the pharmaceutical company, and the prescribing medical staff should share the responsibility and by how much. Existing product liability laws are used to the method of narrowing down design defects or manufacturing defects to a judgment error by a specific subject at a specific time, but in nanomedicines where artificial intelligence changes design variables on its own while learning vast data, it is not easy to trace back exactly when and at which stage that error occurred.

Voices inside and outside the legal field are growing louder, saying that we must first create new rules to share responsibility among the company that made the design algorithm, the hospital that took it and fine-tuned it for a specific patient, and the medical staff who gave the final prescription. It seems that nanomedicine will become a standard in the hospital only when we work together to build explainable artificial intelligence technology that shows how the algorithm made its choice so people can understand it, carefully run cell and animal tests, and bring together the different rules of each country into one.

This trend continues in Korea's pharmaceutical and bio industries as well. Nano Korea, held in 2026, was a global-scale event that brought together nanobio, 3D printing, and precise measuring tools, and target delivery vehicles and diagnostic devices designed by artificial intelligence were treated as major topics. In Korea as well, companies are appearing one by one that use artificial intelligence tools from the very beginning of designing nanoparticles, especially in the fields of new anti-cancer drugs and gene therapy.

Domestic regulatory agencies are also refining their steps while closely watching the changes in the screening standards in the United States and Europe. It seems that domestic pharmaceutical companies, research institutes, and the lawyers who advise them must make an effort to constantly follow overseas clinical data and regulatory trends so they do not fall behind in this flow.

In the winter of 2020, the lipid nanoparticles that first made themselves known to the world in a freezer box at minus 70 degrees have now grown into carriers designed by artificial intelligence that can accurately find deep sores in the brain and the center of tumors. In the meantime, the hands making nanoparticles have largely shifted from the hands of researchers holding test tubes to the calculations of artificial intelligence reading data. Even so, the work for people to do has not disappeared. Researchers, medical staff, and regulatory officials still handle the work of deciding which data to trust, which risks to take, and what to ask the patient first.

The remaining task is to build a bridge that safely carries this detailed blueprint from the laboratory screen to the hospital clinic and pharmacy shelves. The time when that bridge is completed will ultimately be decided by the quality of the data, the speed of the rules, and the judgment of the person who stays by the patient's side to check safety until the very end. It seems that only when those judgments add up one by one will nanomedicine be able to remain as an everyday tool in a normal clinic, rather than a technology that

only shines in moments of crisis like the COVID-19 vaccine.

If we look at Korea, the history of companies directly making and selling nanocarriers is not short either. The anti-cancer drug GenexolPM, which Samyang Biopharm received approval for in Korea in 2007, is a medicine that traps paclitaxel inside a polymer micelle, and it is considered the first polymer micelle anti-cancer drug approved in the world. This means that domestic pharmaceutical companies have already been handling nanocarriers for nearly twenty years. Recently, ST Pharm announced that it will expand its business area into integrated RNA contract development and manufacturing, which includes messenger ribonucleic acid and gene scissors treatments. They plan to do this by using their self-developed lipid nanoparticle formulation technology STLNP and messenger ribonucleic acid capping technology SmartCap, building on their position as the world's third-largest company in oligonucleotide contract development and manufacturing.

Before this, ST Pharm also had the experience of working together with the research team at Ewha Womans University to jointly develop a lipid nanoparticle delivery platform. It looks like domestic nanomedicine technology, which had stayed at the laboratory design stage, has grown into a business aiming for mass production and export.

However, no matter how carefully artificial intelligence designs nanoparticles, there is still one homework problem left unsolved. In 2016, the research team of Professor Warren Chan at the University of Toronto in Canada gathered 117 nanoparticle papers published over the previous ten years and calculated the amount that actually reached the tumor out of the nanoparticles injected into the body. That value was only 0.7 percent based on the median. This meant that if 1,000 nanoparticles were injected, only seven reached the tumor. The research team stated that this number had hardly changed during the ten years before that, too.

Even if a particle is drawn perfectly on a computer, it seems that artificial intelligence has not been able to catch up with its actual journey inside the body, where it flows through blood vessels, gets filtered out by the liver and spleen, and is caught by immune cells. This means that while the design speed has become faster, the old wall of delivery efficiency has not yet been lowered much. Therefore, there is a high chance that the next contest in nanomedicine will be decided by how safely we can deliver the already drawn particles to their destination, rather than by drawing new particles.

The wall of mass production is not easy, either. When nanoparticles, which used to be carefully pulled out a few milliliters at a time with tiny fluidic machines in the laboratory, are moved to factory machines and made in hundreds of liters, the particle size and drug trapping rate often become unstable. The common complaint in the industry is that it is hard to transfer laboratory data exactly to factory standards because the evenness of the lipid nanoparticles breaks even if the speed or temperature flowing through the pipes changes just a little. Artificial intelligence finishes the design in a few days, but the work of turning that design into a safe medicine that people can eat or take without worry still depends on the experience in our fingertips and repeated checks of the process.

PART 4

Preclinical and Clinical Validation and the Precision Medicine Revolution

4.1

Preclinical Innovation: Replacing Animal Testing and Modeling Biology

An IV line was hung on the arm of a patient diagnosed with diffuse gastric cancer. It was a sight that looked no different from the first day of any other clinical trial. The reason this moment was unusual lies elsewhere. The new drug candidate administered to that patient for the first time had never once passed through the body of a mouse or a monkey since its creation. Every process, from design to safety verification, was conducted exclusively within artificial intelligence in computers and organoids on culture dishes.

The medical specialty media outlet BioSpace reported this clinical trial as the world's first case of completing preclinical verification using only artificial intelligence and organoids. Not a single animal came out of the laboratory. Even so, the drug arrived precisely at the patient's bedside.

Until such a scene emerged, the pharmaceutical industry had operated in a different way for a long time. Before a new drug candidate was administered to humans, researchers had verified the absorption, breakdown, and toxicity of the drug by borrowing the bodies of mice, rabbits, and sometimes monkeys. For decades, this method was the sole gateway for new drug development. The problem lies in the fact that the bodies of mice and humans are not built with the same blueprint.

Just as a door will not open if the shape of the lock is different even with the same key, it happened repeatedly that a substance which broke down perfectly well in a mouse's liver reacted completely differently in a human's liver. Due to interspecies differences in gene and protein structures, a significant number of drugs confirmed to be safe in animals ended up causing side effects or failing to show efficacy in actual clinical trials targeting humans.

Rarely is there a case that demonstrates this gap better than what happened in the United Kingdom in 2006. An antibody treatment candidate called TGN1412 showed no particular problems in tests on monkeys even when injected at 500 times the expected human dose. However, when a much smaller amount was actually administered to six healthy volunteers, a cytokine storm - where the immune system goes out of control - erupted throughout their bodies within just a few hours, destroying multiple organs simultaneously.

Although the immune cells of monkeys and humans looked similar on the surface, they were chemically different from each other, and that small difference did not reveal itself on the laboratory bench but only showed its form inside human blood vessels. Following this incident, academia and regulatory authorities painfully accepted the fact that the results of primate tests could not be directly translated to humans.

The industry refers to this gap that opens up between animal test results and actual human responses as the translational research gap. The accumulated results of that gap are clearly revealed in numbers as well. Over 90 percent of new drug development projects end in failure at the clinical stage. Most of these failures are revealed only after the candidate substance enters the human body, which means after a great deal of money and time has already been spent. The time it takes for a single new drug to receive approval averages 10 to 15 years, and the cost incurred reaches 2.6 billion dollars.

The pharmaceutical industry turns Moore's Law backward and calls this paradox Eroom's Law. It is a term pointing to the phenomenon where computer performance improves year by year, while the cost and time to bring a single new drug to the world keep continuously increasing in reverse. This is the reason why long-standing reliance on animal testing has been pointed out as one of the roots of this paradox.

Voices calling for a reduction in animal testing have long been addressed only in the language of bioethics. The argument to reduce the number of animals

sacrificed in laboratories is, of course, justified. However, over the past few years, the center of gravity of this discussion has noticeably shifted. Replacing animal testing is now being treated as a matter of ethics while simultaneously being a matter of increasing scientific accuracy. If models made from human cells and tissues predict human responses much more accurately than a mouse's body, the choice to use fewer animals becomes not only morally right but also a better choice scientifically. As the two reasons of ethical justification and scientific accuracy meet together, the map of preclinical research is being redrawn.

This change is not a story exclusive to large pharmaceutical companies. Building and maintaining animal breeding facilities requires massive initial costs and a long time, but organoid culturing or virtual cell simulations can be started with relatively little capital. A path is opening up for small-scale bio-ventures and university laboratories to jump directly into the preclinical verification stage, which in the past could only be attempted by major pharmaceutical companies. It means that the landscape is shifting into a structure where the novelty of the hypothesis and the quality of the data, rather than the size of capital, determine the outcome.

The first knot of this change was untied in the law. In December 2022, the United States Congress passed the Food and Drug Administration Modernization Act 2.0. This law eliminated the decades-old mandatory provision that animal testing data must be submitted when applying for a new drug approval. The expression used in the legal text also changed from animal testing to nonclinical testing, and this single short phrase became the basis for recognizing human-derived cell models, organoids, organs-on-chips, and artificial intelligence-based computer simulations as legally valid alternative methods. The entire pharmaceutical industry was abuzz. Animal testing became no longer the sole correct answer, but one of multiple options.

Even after the law opened the door, the task of actually widening that door had to be taken on by the United States Food and Drug Administration. The Food and Drug Administration first announced its direction to reduce animal testing in 2024, and in April 2025, it developed this into a concrete roadmap and presented a goal to make animal testing not the standard but an exception in new drug development within the next 3 to 5 years.

The Food and Drug Administration plans to first gradually reduce animal testing for monoclonal antibodies - that is, antibody treatments cloned from a single cell and sharing the exact same shape - and other biological products, before expanding that scope to chemical synthetic drugs. In early 2025, it also released a draft guidance defining how to handle artificial intelligence in the review of drugs and biological products.

A year after releasing its roadmap, the Food and Drug Administration announced that it had achieved its first-year goals. It released a draft guideline aimed at reducing or eliminating primate testing, such as with monkeys, in the development of antibody therapeutics. It also updated guidance on replacing endotoxin test material extracted from horseshoe crab blood with synthetic substances. Simply by changing this one test that had relied on horseshoe crab blood, more than a million animals per year no longer need to be placed on experimental tables. The U.S. House of Representatives has already passed a follow-up bill asking not to stop there but to accelerate the pace of implementation.

The core of the draft artificial intelligence guidance released in early 2025 is the idea of assessing the reliability of artificial intelligence models differently depending on their risk level. Even with the same artificial intelligence, a model used to initially screen new drug candidates and a model used to make final safety judgments without clinical trials are treated with completely different weights. The greater the impact on regulatory decisions, the more carefully one must demonstrate what data the model was trained on and what

procedures are in place to filter it when predictions are wrong. This risk-based assessment framework shows in advance what gates virtual cells or systems biology models will need to pass if they are to actually step into clinical trial design.

The force driving this flow does not come from a single technology but from two different directions together. On one side is in vitro technology - that is, the technology of actually growing human cells and tissues outside the body. On the other side is in silico technology - that is, the technology that takes place inside a computer, moving human cells wholesale into a computer. If organoids and organ-on-chip systems mimic the human body physically in front, then artificial intelligence-generated virtual cells and systems biology models mimic the human body digitally in the back.

The two technologies interweave like warp and weft threads, filling in what each has missed. Predictions from the computer are verified with living human cells, and experimental results in culture dishes are used in turn to refine the computer model more precisely.

There is a reason why all of this is happening concentrated in the mid-2020s. Computational power to train artificial intelligence models has grown to a scale that was unimaginable in the past as graphics processing unit performance improves significantly year after year. At the same time, as the price of experimental equipment that reads genetic activity at the single-cell level has dropped and processing speed has increased, sufficient biological data to train virtual cells has finally begun to accumulate.

With cell culture and microfabrication techniques for making organoids and organ-on-chip systems maturing, experimental tools to physically verify predictions from digital models were equipped at the same time. If three conditions - computational power, biological data, and experimental tools to validate it - had not been equipped together at a similar time, even if organoids and organ-on-chip systems had become more sophisticated, the digital

counterparts to support them would not have emerged.

A virtual cell is a digital twin - a complete replica of a single cell transferred into a computer. Researchers stack data obtained at multiple levels, such as single-cell transcriptomics, spatial transcriptomics, and proteomics, to mathematically reproduce the cell's gene expression network and metabolic pathways. Single-cell transcriptomics reads out which genes each cell is using and how much; spatial transcriptomics maps out exactly where in the tissue that activity is taking place; and proteomics is the result of measuring how much protein, ultimately produced by genetic information, exists within the cell.

It is similar to running thousands of simulated flights in a computer before actually flying an airplane. By administering drug candidates to this virtual cell or conducting in silico experiments where specific genes are artificially manipulated, researchers can anticipate how cells will respond and whether they might attach to unintended targets and cause side effects, all without actual culture dishes or animals.

For virtual cells to be useful, they must be supported by systems biology modeling that calculates everything from the movement of individual molecules to changes in entire tissues. This modeling differs from machine learning, which statistically sweeps through data and finds only patterns. It uses biophysical mechanistic models that solve with differential equations how certain causes lead to certain effects, alongside artificial intelligence. Platforms like COPASI or CellNOpt reproduce the process of signal transmission and metabolism within cells, while multicellular simulators like PhysiCell or CompuCell3D handle the behavior created by multiple cells interacting, such as tumors growing or immune cells infiltrating tissues.

Only after this chain of calculations - from molecules to cells to tissues - is in place do virtual cells finally become reliable prediction tools.

Recursion Pharmaceuticals can be cited as an example that actually implemented this approach in industry. This company combined laboratory robot automation equipment with a machine learning platform called inVivoPRINT to transform the very method of preclinical research. While robots photograph cells and micro-tissues 24 hours a day and sensors measure numerical values, deep learning models read this vast imagery and data in real time to extract the phenotypic signals of toxicity and efficacy shown by cells exposed to the drug.

By comparing these compound fingerprint-like signals obtained this way to databases of known toxic substances, risks of fatal organ damage that would have emerged only in late development can now be caught early in the research phase. Thanks to filtering out dangerous candidates early, Recursion Pharmaceuticals brought a drug candidate all the way to the human clinical trial stage just 18 months after starting the project.

Recursion Pharmaceuticals is not the only one walking this path. So-called autonomous laboratories, in which robots design and execute experiments and decide on the next experiments themselves without human intervention, are growing together at multiple pharmaceutical companies and research institutions. While human researchers can hardly process more than dozens of experimental dishes per day no matter how hard they try, an autonomous laboratory with robot arms and artificial intelligence working in tandem tests thousands of combinations in the same time and immediately reflects the results in the design of the next experiment.

Thanks to this laboratory that runs without resting all night, the speed of selecting new drug candidates has become incomparably faster than the days when humans tested them one by one by hand.

Virtual cell technology took several steps forward at once between 2025 and 2026. The Arc Institute in the United States unveiled State3, a first-generation artificial intelligence virtual cell model, in mid-2025. This model was created by

learning from 170 million pieces of cell data and over 100 million single-cell perturbation response data obtained from more than 70 cell lines. The Arc Institute did not stop there and held a public competition called the Virtual Cell Challenge, where multiple artificial intelligence models competed to see how closely they matched actual biological experimental results.

Over 1,200 teams from 14 countries flocked to submit more than 300 final results. Academia began calling this competition the Turing test for virtual cells. Just like the original Turing test, where a computer passes the test when its conversation cannot be distinguished from a conversation with a human, this means it is a test to compete on whether the predictions produced by the virtual cell are accurate enough to be indistinguishable from the results of experiments on actual living cells.

The movements of giant technology companies also coincided around the same time. The Chan Zuckerberg Initiative released for the first time a universal digital representation of cells and cell states directly learned from biological data, so-called artificial intelligence virtual cells. Soon after, they partnered with Nvidia to open a Virtual Cell platform, laying the foundation to handle the data needed to build and grow virtual cell models at scale.

This foundation also released an inference model called rBio, which is a tool that finds answers based on virtual cell simulations when researchers ask biological questions in plain sentences. Google DeepMind also announced that it is working on its own virtual cell project. Virtual cells have now become a stage of competition where several giant technology companies jump in simultaneously, rather than a gamble by a single company.

No matter how sophisticated the virtual cells inside a computer become, they cannot replace the three-dimensional physiological phenomena of a living human body exactly as they are. Therefore, a physical counterpart is needed to actually verify the predictions. Organoids take on that role. Since they are made of human cells, not mouse cells, it is like a native speaker telling you

directly without going through an interpreter. An organoid is a miniature organ made by taking adult stem cells obtained from a patient or induced pluripotent stem cells, which are fully grown skin cells turned back to their childhood pluripotent state, and growing them three-dimensionally on a dish so that they have a cellular composition and structure similar to an actual organ, as well as living functions, albeit limited.

Organoids show the unique cell-to-cell signaling and tissue microenvironment of humans exactly as they are, which flat two-dimensional culture dishes or animal bodies cannot mimic. Organoids are rapidly expanding their uses as a tool to check in advance inside the laboratory how well new drug candidates actually work on humans.

In the face of diseases where nerves slowly collapse, such as Alzheimer's disease and Parkinson's disease, organoids show exceptionally great usefulness. The brains of rodents, including mice, cannot properly mimic the wrinkled cortical structure unique to the human brain, the complex glial cell network, or the pathological phenomenon of proteins clumping together. That is why for a long time, rodent brains have been called by the dishonorable name of the graveyard for new drug development in the central nervous system. Mini brains and midbrain organoids made from human cells are different. They recreate in three dimensions the core pathology of Parkinson's disease, the degeneration of dopamine-producing neurons or the abnormal clumping of a protein called alpha-synuclein, very similarly to an actual patient's brain.

Recently, an experimental model has even emerged that precisely controls the clumping process of alpha-synuclein by combining optogenetics technology, which controls cell activity with light, with midbrain organoids. On this model, researchers are directly confirming the effects of new drug candidates that protect nerves within a human biological context.

The weight of these technologies becomes even greater in the development of treatments for rare diseases. There are quite a few diseases where even animal models are not properly equipped due to the small number of patients, and in such cases, organoids cultured directly from a single patient's cells often become practically the only means of preclinical verification. Instead of pouring years into creating and verifying new animal models, one can start from a single tissue sample of a patient, grow an organoid within a few weeks, and test the drug response. This time difference of a few weeks is crucial for the families of patients with rare diseases.

Organoids also have gaps. They lack a vascular network, miss interactions with the immune system, and cannot capture the physical forces that tissues inside the body actually experience. To fill these gaps, organ-on-a-chip devices combining microfluidics, namely microphysiological systems (MPS), have stepped to the front lines of preclinical research. In this field, in addition to Emulate, companies like Mimetas, TissUse, and Hesperos have jumped in, each targeting different organs. An organ-on-a-chip engraves fluid channels thinner than a hair inside a transparent polymer chip the size of a thumb, and grows human-derived tissue cells and vascular endothelial cells together inside them.

This single small chip replicates the mechanical environment occurring inside the human body exactly as it is, down to the pressure created by flowing blood, the squirming movement of the intestines, and the rhythm of the lungs inflating and deflating. What the organ-on-a-chip recreates is not just the shape of the cells, but the physical conditions those cells actually experience inside the body, namely the entire pressure, flow, and movement.

Each individual organ-on-a-chip does not stop at mimicking a single specific organ like the liver, heart, kidneys, or lungs. By connecting multiple organ chips with thin tubes, one can continuously observe outside the culture dish the entire process from when a drug is absorbed into the body, metabolized,

and excreted out of the body, namely the entire pharmacokinetics. For example, if a gut chip and a liver chip are connected side by side, it can recreate the sequence exactly as it occurs in the actual body, down to the order where an oral drug is absorbed in the intestines and then immediately passes to the liver to be broken down.

This connection order between organs is information that cannot be obtained from individual organ chips alone. A blood-brain barrier chip co-cultured with brain vascular endothelial cells, astrocytes, and pericytes allows AI-designed brain disease therapeutics to penetrate the barrier surrounding the brain more accurately than animal testing. Multiple microelectrodes embedded on organ chips made from human cells detect in real-time whether drugs damage the liver or cause heart rhythm abnormalities.

These electrodes read subtle changes in electrical signals emitted by cells in units of seconds. For heart cells, they immediately detect the moment when the beating rhythm becomes irregular; for liver cells, they immediately capture signals leaking as the cell membrane is damaged. When predictions from virtual cells in a computer are uncertain, data from this physical chip removes that uncertainty. This cycle is completed when *in silico* points the direction first and *in vitro* confirms whether that direction is correct.

There is a reason why the liver has particularly been a bottleneck in new drug development. Most drugs are broken down through the liver, and in this process, unexpected byproducts are often created that damage liver cells. Drug-induced liver damage has been cited as a common reason for drugs being withdrawn from the market after approval, and it was not rare for substances that were fine in animal testing to cause problems only in human livers.

It is only recently that this technology has gained force in the regulatory field, crossing the laboratory fence. The U.S. Food and Drug Administration solidified the IStand system, a pathway for reviewing innovative scientific and

technological approaches, as an official system in mid-2025. Through this pathway, Emulate's LiverChip is taking the final steps to be recognized as a tool for detecting drug-induced liver damage in advance. The process involves being validated for predicting actual liver damage in human livers based only on morphological changes in tissues exposed to toxic substances and fluctuations in albumin and ALT protein levels. Once this recognition is obtained, the LiverChip becomes not a tool used by one specific company but an officially recognized tool that all new drug development programs can use within defined purposes.

Here, the defined purpose, or context of use, is regulatory language that pins down exactly what questions this chip can answer. Being recognized for the purpose of predicting liver toxicity does not automatically grant recognition for cardiac toxicity or kidney toxicity. We are approaching the first case where data generated by an organ chip is recognized as official evidence in regulatory review, beyond reference material for individual experiments.

Organoids are also moving out of the laboratory to patients' bedsides. Between 2023 and 2025 alone, 139 clinical trials using patient-derived organoids were registered, and this trend continues into 2026. A Phase 2 clinical trial using organoids to test drug response in advance and determine treatment direction has begun in early 2026 for patients with unresectable hepatocellular carcinoma. Even with the same hepatocellular carcinoma diagnosis, the genetic characteristics of tumors vary by patient. This clinical trial examines on a dish how organoids derived from each patient's tumor respond to various drugs, and then prescribes the drug that works well for that person's organoid.

The order itself differs from the previous trial-and-error method of prescribing, where you try a drug and switch to the next one if it doesn't work. This means that organoids no longer remain tools for selecting new drug candidates in the laboratory but are now already being used to decide which of the

already-approved drugs to use for a specific patient.

Korea's activities are also not sidestepped here. The Ministry of Food and Drug Safety and its subsidiary Korea Center for the Validation of Alternative Methods are leading efforts to raise non-animal test methods such as organoids, organ chips, and human tissue models to international standards, and have set a goal to be listed as internationally recognized test methods of the Organization for Economic Cooperation and Development by 2028.

Domestically, startups like OrganoScience and NextAnd Bio are also riding this trend by commercializing organoid-based new drug evaluation platforms.

At the Korea FDA Regulatory Science Society Spring Academic Conference in 2026, a special session covering organoid-based animal alternative test methods was held separately, where domestic researchers and regulatory authorities are working together to refine these standards.

Despite all this progress, many knots remain unsolved. Artificial intelligence models tangled in tens of millions of variables often cannot explain in human-understandable language why they made a judgment that a particular compound is dangerous, even when they have judged it so. This is no different from a student who gets the answer right but does not show the solving process. What regulatory authorities require when approving new drugs is not just the result but the logic leading to that result, so predictions that cannot explain the basis, no matter how accurate, have difficulty clearing the review threshold.

Models that showed high accuracy with compounds similar to the data used in training often lose predictive power suddenly when encountering completely unfamiliar new substances or biological data from different races. Recent research also points out that due to data leakage where information corresponding to the answer is accidentally mixed into the training process, falsely positive results that appear much more accurate than reality are produced.

There are also problems outside the technology. To exchange virtual cell models created by different research institutions or companies and validate them again, there needs to be a common language, but standard model technical languages like SBML or CellML have not spread widely, leaving data and models scattered here and there. To properly train virtual cells requires genetic information and medical records from hundreds of thousands of people, which directly conflicts with Europe's personal information protection laws or the United States' health information protection laws.

Alternatives such as federated learning, where raw data is not exported outside the institution and only encrypted results are exchanged, are being discussed, but the systems and legal provisions to support this have not yet been put in order in many countries.

The organoid cell lines and biological data accumulated so far are concentrated in specific regions and populations, so it is difficult to say with confidence how accurate a model trained on this data will be when applied directly to patients with different genetic backgrounds. When a judgment made together by artificial intelligence and microphysiological systems causes unexpected harm to a real patient, there are also no legal standards yet to determine whether the company that made the software, the hospital that provided the data, or the pharmaceutical company that made the final decision should bear the responsibility.

Nevertheless, the IV line inserted into the arm of that gastric cancer patient has already begun to tell a different story. A drug that passed through neither mice nor monkeys has reached a human vein. In the past, this is a journey where just the animal testing phase would have taken several years. No one can yet be sure how many patients this drug will save in the years to come. The clinical trial has only just taken its first step. However, the trend of computer servers and culture dishes stepping up as the first gateway to new drug development, replacing the mouse cages in the laboratory, has already reached a point

where it is difficult to reverse.

In Korea, there have also been continuous attempts to create new laws in line with this trend. Until now, Representative Nam In-soon and Representative Han Jeoung-ae each proposed bills to enact the Act on Alternative to Animal Testing, but due to differences in position among the Ministry of Agriculture, Food and Rural Affairs, the Ministry of Climate, Energy and Environment, the Ministry of Food and Drug Safety, and the Rural Development Administration, both bills failed to pass the threshold of the National Assembly and disappeared. In December 2025, Representative Song Ok-joo proposed this bill again as the main sponsor. Since this is a proposal brought forward after the relevant ministries aligned on the content in advance, there are assessments that the situation is different from the time of Representatives Nam In-soon and Han Jeoung-ae.

The bill includes provisions for government ministries to form a consultative body together and establish a basic plan, to set up centers to verify alternative animal testing methods by field, and to investigate the actual conditions every 5 years. There are also numbers that were revealed along with the introduction of the bill. In the single year of 2024, the number of laboratory animals used domestically reached 4.59 million, and more than half of these experiments were classified into the highest severity category, causing extreme pain by exposing them to toxic substances or inducing tumors without anesthetics or painkillers.

Blood vessels and the immune system are not the only gaps left in organoids. Midbrain organoids show the movements of individual nerve cells producing dopamine similarly to a human brain, but they have not yet caught up to the circuits where those nerve cells extend long axons to other brain regions to send and receive signals. The human brain operates as a single circuit with the cortex, hippocampus, and basal ganglia sending and receiving signals with one another, but such long-distance connections are missing in a single organoid

grown alone on a dish.

Recently, studies have been emerging that physically connect several organoids to create structures called assembloids, and use electrodes to measure the neural signals traveling between them. The direction is right, but it is at a toddler stage, and phenomena that only emerge when the entire brain moves together, such as storing memories or evoking emotions, cannot be observed on a dish in the first place. Reproducing the complex circuits at the level of a human brain in their entirety remains a distant goal.

It seems that the demand to reduce animal testing and the demand to securely protect human safety will not easily reconcile in the future. The incident where TGN1412 appeared safe in the bodies of monkeys but caused the immune system to spiral out of control in human bodies also demonstrates the fact that there are risks that a single organoid or organ-on-a-chip cannot yet completely prevent. The responses created by the entire immune system sending and receiving signals with one another, or the chronic toxicity experienced by multiple organs simultaneously, remain areas that current non-animal testing methods find difficult to catch up with.

Until enough evidence accumulates to confidently filter out even these risks, rather than banning animal testing all at once by law, taking them over one by one starting from lower-risk areas is highly likely to be the more realistic path.

4.2

Digital Twins and the Modernization of Next-Generation Clinical Trials

On a monitor in a conference room at the U.S. Food and Drug Administration, a human heart was beating in red. It was not a real heart, but a digital replica sent by Dassault Systèmes, a software company headquartered in France - the so-called Living Heart. This project, which began in 2014, completed the world's first three-dimensional digital heart replicated exactly from a human heart, and later formed a five-year joint research agreement with the Food and Drug Administration to be adopted as an official virtual testing environment for validating cardiovascular medical devices. Around the same time, in another conference room with documents before them, a proposal submitted by the artificial intelligence startup Unlearn.ai was in the hands of the reviewers.

The proposal was to fill the patient group that did not receive the new drug - in other words, the control group - with virtual patient data calculated by a computer instead of real people. At the official Type C meeting where new drug developers discuss clinical trial design with regulatory authorities in advance, this proposal was accepted. It was the first moment when regulatory authorities officially acknowledged that patients created by a computer, not by actual people, could fill half of a clinical trial.

Traditional clinical trials remain the final gate for new drug approval today, but the path through that gate is growing increasingly difficult. More than 80 percent of clinical trials worldwide fail to recruit patients on time, extending timelines, and the pharmaceutical industry as a whole spends over 200 billion dollars in research and development per year. Even with all this effort, fewer than 12 percent of new drugs make it to final approval. Patients in the control

group who do not receive the new drug and receive only a placebo - that is, a fake drug - miss their best opportunity for treatment. The structure of randomized controlled clinical trials, which requires following thousands of people for years to produce statistically meaningful results, places a heavy burden on both patients and budgets.

In attempts to reduce this burden, digital twin technology and clinical trial simulation technology - which redesign clinical trials by weaving together artificial intelligence and vast real-world data - are emerging as the driving force to complete precision medicine. Reducing the time and cost required to enroll and follow a single patient is no longer a choice for a few companies but is approaching a survival condition for the entire new drug development industry. Ultimately, the purpose of this change is not merely to bring a new drug to market faster, but to ensure that the new drug reaches the patient who truly needs it in the way that is truly necessary.

Before a new pilot climbs into the real sky, they first sit in a flight simulator - in other words, a simulation device. Although the sky on the simulator screen is fake, the engine sound, the vibration of the instruments, and the feeling of being pushed by a sudden gust are designed to feel almost identical to real flight, so the pilot can practice making mistakes before carrying real passengers. Digital twin technology creates this simulator for each individual person. When a patient's genetic information, hospital medical records, the heart rate and sleep measured by a smartwatch on their wrist, the results of imaging tests, and even protein levels in the blood are gathered inside a computer, a digital replica that moves in almost the same way as that person is created.

Instead of giving an unverified new drug to real patients first, doctors and researchers first test the drug on this digital replica. Only after confirming in the computer what kind of reaction occurs when the dose is increased and whether side effects start in the liver or the heart do they proceed to actual

prescription. Just as making dozens of mistakes inside a simulator does not injure real passengers, only after testing the new drug hundreds of times inside the digital replica does it approach real patients.

Similar attempts have been made in the past. However, past models only calculated how a particular drug was absorbed and eliminated on average across a population as a whole and were not connected to real-time data from individual patients. They remained at the level of calculators confirming static hypotheses. With the recent arrival of deep learning, this calculator has moved into an entirely different dimension. The current technology, called an *in silico* twin because it takes place inside a computer, accepts wearable device data and electronic health records, high-resolution imaging, and multi-omics data that examine genes and proteins together in real time 24 hours a day, and it learns on its own how disease progresses inside a patient's body and continues to update itself.

Because a single sensor produces tens of thousands of data points in one day, markers that used to be measured only once when going to a hospital can now be checked at almost every moment. This means that the digital twin from yesterday and the digital twin from today can produce different predictions, and it also means that it follows the actual patient's physical condition more closely.

In fields like precision oncology, where each tumor has different characteristics, identifying patients with high likelihood of responding to a specific targeted therapy is crucial to the success or failure of a clinical trial. In times before such stratification existed, patients who responded to the drug and those who did not were tested together, and it was not uncommon for truly effective new drugs to be deemed ineffective, buried in average values. Artificial intelligence analyzes genetic and pathological data accumulated in digital twins to identify a patient subgroup with a high likelihood of responding to a drug - that is, patients carrying specific biomarkers, substances that signal

changes in the body.

Thanks to the precision of examining even a single gene variant or a single protein level, a drug that works in only one out of ten people can be precisely identified in that one person and lead a clinical trial to success. As a result, clinical trial designers can greatly reduce the number of patients needed while making the statistical confidence required to demonstrate drug efficacy even stronger. This is where the paradox emerges: obtaining more reliable results while recruiting fewer patients.

Opposite to this task of selecting patients, the task of creating a control group faces it like a mirror. In the real world, artificial intelligence filters out patients who will respond to the new drug, but in the digital world, the same artificial intelligence creates a virtual control group that is statistically twin to the patient. The materials are real-world data (RWD) and real-world evidence (RWE) that already exist in the world, such as clinical trial records accumulated in various hospitals in the past, global disease registries, and electronic health records. Based on this data, researchers recreate in the computer how the disease progresses when the new drug is not used at all, and compare the results side by side with the progress of patients who actually received the new drug.

The front-end task of setting up a treatment group by filtering real patient data with artificial intelligence, and the back-end task of setting up a virtual control group by filtering real-world data with artificial intelligence, is a structure that repeats the same principle on both sides. The fact that the eyes selecting patients and the eyes creating the control group are ultimately the same artificial intelligence means that an era has opened where the real world and the digital world must be calculated simultaneously whenever a clinical trial is designed.

As digital twins enter clinical trials, pharmaceutical companies are now able to estimate the probability of success of a single clinical trial long before

recruiting actual patients. The process of first matching the mechanism of action of the new drug under development and the digital twin data of the target patient group in a computer, and filtering out designs that are likely to fail even before the clinical trial begins, is akin to a kind of virtual dry run. Because years of time and massive costs simply disappear if a single phase 3 clinical trial ends in failure, each failure filtered out by this dry run leads to the effect of saving considerable resources for the industry as a whole.

Although success in a computer does not guarantee success in real patients as is, simply filtering out designs with a low probability from the beginning is helpful enough to raise the success rate of new drug development. Because artificial intelligence is involved from the stage of designing clinical trials, there is also the advantage of being able to fix design flaws in advance that would have been revealed only after patient recruitment ended in the past.

The Prognostic Covariate Adjustment (PROCOPA) technique developed by Unlearn.AI is a method of reducing the size of the control group by applying the prognostic score of an individual patient predicted by a digital twin to a statistical model. Unlike the previous statistical method that only compared averages without reflecting the individual patient's prognosis, this technique first factors in the different recovery speeds and risk factors of each patient and then singles out only the pure effect of the new drug. The European Medicines Agency (EMA) issued an opinion qualifying this technique for use in phase 2 and phase 3 clinical trials, and it was the first time that European regulatory authorities raised their hands for an artificial intelligence-based methodology in this way.

The United States Food and Drug Administration also expressed positive opinions on the same technique in multiple Type C meetings. According to the figures revealed by Unlearn.AI, the variance of the treatment effect estimate can be reduced by about 10 percent to 20 percent, which leads to the result of reducing the number of patients needed for the control group by 20 percent to

30 percent, and at most even more. Hybrid clinical trials, which use actual clinical trial data and virtual data created by digital twins together, are now settling as a standard rather than an exception.

In addition to Dassault Systemes and Unlearn.AI, companies such as Nova In Silico and Deep Intelligent Pharma have jumped into the same market, expanding the competition further.

The industry has begun to call this design that combines half of the actual clinical trial and the other half created by digital twins as TwinRCT. Just like the name, it is a method of seeking answers half from humans and half from their digital twins. Unlearn.AI is submitting this name to regulatory authorities as the official name of its technology, and such naming is also an attempt to translate a still unfamiliar concept into the language of the clinical field.

TwinRCT is now gradually spreading beyond a few rare disease studies to the area of common chronic diseases. When designing a single clinical trial, how much to divide the proportion of humans and digital twins seems to be a question that clinical trial designers will newly face in the future.

In the field of rare diseases, the usefulness of this technology stands out even more. If the number of patients itself is limited to hundreds in the entire country, there is no luxury to divide them in half into a placebo group and a treatment group in the first place, and dividing them like that itself becomes a statistically unreasonable demand. In such diseases, the synthetic control group created by digital twins is considered practically the only realistic alternative. This is even more so in the area of pediatric rare diseases, where it is difficult for parents to readily agree to assign their children to the placebo group.

Considering these circumstances, new drug systems for rare diseases in various countries often apply control group requirements flexibly. The method of later attaching a comparison group created by digital twins to a single-arm

clinical trial where only the new drug is administered to all clinical trial participants is also a growing trend. Industry journals are also stating that the development of new drugs for rare diseases is facing a new turning point through digital twins and virtual clinical trials.

Setting aside a separate placebo group for these patients is often difficult to justify ethically. Giving only fake medicine to patients whose disease progresses quickly and making them wait for years is a choice that is even harder to accept if there is a possibility that the new drug will work. The synthetic control group created by digital twins opens up an opportunity for these patients to receive the new drug instead of a placebo, while still allowing the comparison data required by the clinical trial to be secured as is. This awareness of the problem itself is not new. Even before the emergence of artificial intelligence, historical control groups created only from past records were frequently embroiled in reliability controversies, and digital twins are closer to an attempt to supplement that old debate with much more sophisticated statistics.

That said, synthetic control groups cannot be used indiscriminately. Virtual control groups created from past data carry the risk of already differing from today's clinical environment, and if standard treatment has changed over a few years, old data is difficult to use as a comparison standard for today. For this reason, the U.S. Food and Drug Administration clarified in its 2023 guidelines that external control groups would be reviewed on a case-by-case basis, and narrowed their use to cases of severe diseases lacking other suitable treatments where objective measurable evaluation indicators exist.

Adaptive clinical trials are a method in which patient response data coming in during the trial is reviewed and dosing or target patient populations are adjusted as needed. Reinforcement learning algorithms read interim analysis results, increasing doses for certain patients and excluding patients showing no response or at high risk of adverse effects from the trial early. A preliminary

clinical trial of CURATE.AI targeting solid tumor patients is a concrete example of this method in action. In standard chemotherapy regimens for gastrointestinal cancer patients such as XELOX or XELIRI, the dose of the oral anticancer drug Capecitabine contained within was continuously adjusted according to individual patient response data.

This approach of treating a single patient as an independent clinical trial is called N-of-1, and it differs from conventional prescribing in the order itself - instead of fitting patients to standard dose charts, it finds doses suited to each patient in real time. Results of this preliminary clinical trial were published in 2025 in the international journal *npj Precision Oncology*, confirming that dose adjustments tailored to each individual patient are actually feasible.

The draft guidance released by the U.S. Food and Drug Administration in January 2025 adopted a risk-based approach. Rather than blindly trusting or blindly doubting each artificial intelligence model, it adjusts the level of scrutiny based on how and where the model will be used. If a digital twin is used only as reference material for determining initial doses, relatively light validation is sufficient, but if used as a synthetic control group that completely replaces placebo, it must pass far stricter validation. Before submitting a model, a new drug developer must first prepare documentation describing what data and structure the model was built from, whether it performs the same on new patients not used in training, and then consult with regulatory authorities about the model.

It is a cumbersome process, but conversely, it also opens a path for evidence created by artificial intelligence to be reviewed on equal footing with evidence created by humans. This approach is not much different from how automotive or aircraft parts are certified by assigning different safety grades to each component, so it is not an entirely unfamiliar framework even from the regulatory authorities' perspective.

That this trend does not remain confined to experiments at a handful of companies is confirmed by other clinical trials underway in various locations. At Cleveland Clinic in the United States, a randomized controlled clinical trial is underway to verify the effectiveness of full-body digital twin-based precision medicine targeting type 2 diabetes patients, with results expected by 2026. In April 2025, a separate study also began to validate the digital twin-based clinical research system itself. Market research organizations project that the global market size for digital twins will grow from approximately \$17.7 billion in 2024 to \$259.3 billion by 2032, at a rate close to 40 percent annually.

Large pharmaceutical companies including French pharmaceutical company Sanofi have begun publicly introducing, through their own media, the work of redesigning clinical trials using artificial intelligence-created digital twins. As platforms like Outcomes4Me, which help patients directly check their own treatment progress, have emerged, digital twins are expanding beyond researchers' tools to become tools used by patients themselves.

If synthetic control groups reduced the number of people required to serve as control participants, wearable devices further reduce hospital visits even for remaining study participants. When patients wear devices such as smartwatches and smart patches, digital biomarkers such as heart rate, sleep, and activity levels accumulate 24 hours a day without patients needing to come to the hospital. These decentralized clinical trials (DCT), which are not tied to the physical location of a hospital, are estimated to exceed a global market size of \$10 billion as of 2026.

The work of the U.S. government's Biomedical Advanced Research and Development Authority (BARDA) partnering with large pharmacy chains like Walgreens to establish neighborhood pharmacies as clinical trial sites is also part of the same effort to lower barriers to hospital access. The Food and Drug Administration also opened a new Clinical Trial Innovation Center in April 2024 and finalized guidelines for decentralized clinical trials in September of the

same year, adding regulatory support to this trend.

This flood of data is read by large language models. Artificial intelligence reads unstructured clinical records and patient-written text in real time, detecting adverse drug reactions or drug-drug interactions with sensitivity exceeding 90 percent. There have been research teams that automatically extracted adverse reactions to influenza vaccines from reports, and there have also been research teams that used large language models applying Tree of Thoughts reasoning methods to automatically identify cardiovascular adverse reactions in clinical records.

Rural patients and minority patients who find it difficult to travel to hospitals can also participate in clinical trials through this method, so decentralized clinical trials are solving not only efficiency issues but also equity issues in clinical data. The advantage of this automatic detection over patients directly entering their own symptoms is that it does not miss early signals that patients themselves do not report because they consider them insignificant.

In the second half of 2024, Dassault Systèmes, together with the Food and Drug Administration, published the world's first in silico clinical trial guide based on V&V 40, a validation standard of the American Society of Mechanical Engineers. This project, which began with the combined efforts of doctors, engineers, and government officials, entered beta testing in February 2025 for the next generation Living Heart model, which can be finely adjusted for each individual patient or specific patient group, Dassault Systèmes stated. Now that the Living Brain has followed the Living Heart, the same method is currently expanding to almost all human organs, including the liver and lungs.

The Living Heart project was also named on the 2025 list of ideas changing the world by the American business magazine Fast Company. The international academic journal Nature Medicine published a paper in 2026 stating that digital twins and in silico clinical trials had entered a turning point in new drug development, and research on augmenting clinical trials with digital twins was

published in the academic journal npj Systems Biology around the same time.

Behind these brilliant achievements, heavy obstacles remain. The first problem encountered is the so-called black box problem, where the inside of the judgments made by artificial intelligence cannot be seen. Even if a deep learning model intertwined with tens of millions of parameters concludes that a specific treatment is effective or warns that liver toxicity will occur, if it cannot clearly explain to doctors and regulatory authorities what biological pathway it went through to reach that judgment, it is difficult for it to be adopted as key evidence in new drug approval reviews.

This is because a computer cannot be left to make decisions where a patient's life is at stake in a state where no one can explain why it judged that way. Among doctors, there is still a strong attitude that they will sign a prescription only after understanding the reason, rather than simply following the results provided by a computer. Therefore, explainable artificial intelligence techniques that show the final predicted value alongside a few factors that greatly influenced that prediction are recently emerging as a realistic alternative.

The limitations of training data also hold things back. Models trained on data skewed toward specific races or specific disease patterns experience a significant drop in prediction accuracy when they encounter unfamiliar new drug candidates or different demographic groups, and cases where that accuracy dropped to a level of 0.3 have been reported in academia. The data leakage problem, where clinical result information that should appear later accidentally mixes into the input data during the training process and inflates the model's performance indicators beyond reality, is also repeatedly pointed out. These errors often do not reveal themselves during the clinical trial stage, but pose a risk of appearing in actual prescription settings as false positive judgments that incorrectly announce a risk that does not exist.

For these reasons, regulatory authorities have recently started demanding that new drug submission data include out-of-sample validation results, which separately confirm performance on a new patient group not used at all for model training. Thus, academia and the pharmaceutical industry are demanding strict validation procedures, covering everything from software unit tests to validation of agreement with biological responses, as an essential gateway for new drug development.

To create a sophisticated digital twin, highly sensitive personal information like genomic information and electronic medical records must be collected in massive quantities. Under the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe, the very act of moving data individually held by hospitals and pharmaceutical companies to external servers to gather it in one place is effectively blocked. What has emerged as an alternative is federated learning and differential privacy technologies, where the patient's original data remains exactly as it is on the hospital's internal servers, and only encrypted model weights are exchanged. In multinational clinical trials that cross borders, it is also common for a data processing method that is legal in one country to be illegal in another, so the burden on legal staff designing international joint research is growing ever heavier.

The cost of standardizing different hospital data formats into one and maintaining an ironclad level of cybersecurity remains considerable. Therefore, large hospitals are also trying an approach where, instead of directly exchanging each other's original data, they only use data that has completed anonymization verification through a trusted third-party institution for research.

Concerns are deepening within the pharmaceutical industry over how to document the ALCOA++ principles, which demand data integrity - meaning standards like attribution tracing who created the data and when, along with

accuracy, completeness, and consistency - inside artificial intelligence algorithms. In addition to this, when a new drug approved based on virtual patients and synthetic control groups designed by artificial intelligence causes unexpected side effects after going on the market, the legal standards organizing who should be held responsible among the company that made the software, the hospital that provided the data, the doctor who issued the prescription, and the regulatory authority that stamped the approval have not yet been properly established in any country.

For now, new drug developers are merely filling this legal gap by explicitly dividing responsibility in advance within contracts, and there is a high possibility that this uncertainty will continue for the time being until clear standards are set by court precedents or separate legislation. Although governments in various countries are reviewing related legislation, it is still difficult to find a country that has moved beyond the bill discussion stage and entered actual implementation.

The United States Congress passed the FDA Modernization Act 2.0 across 2022 and 2023, removing the mandatory animal testing clause that had been a prerequisite for new drug approval for decades and officially recognizing artificial intelligence-based in silico simulation data as an alternative means to prove safety and efficacy. In January 2025, the Food and Drug Administration released a draft guidance to be followed when using artificial intelligence for regulatory decisions on drugs and biological products, demanding that the model's purpose be defined in advance and that data management, model structure, and performance on data not used for training all be validated.

In December of the same year, it finalized real-world evidence-based medical device review guidelines and announced that it would accept anonymized large-scale medical databases as evidence material even without raw data that could identify individual patients. The European Medicines Agency adopted a reflection document on artificial intelligence use across the entire

pharmaceutical product lifecycle in September 2024, and recommended to the pharmaceutical industry that it transparently explain the model structure and that humans take responsibility and provide oversight to the end for results. In July 2025, it submitted a separate guidance draft on external control groups for public consultation. The United Kingdom Medicines and Healthcare Products Regulatory Agency (MHRA) also released a guidance draft on external control groups based on real-world data in May 2025.

Korea's Ministry of Food and Drug Safety also established guidelines containing recognition criteria for clinical trial data collected by digital devices in December 2025. In January 2026, the Food and Drug Administration and the European Medicines Agency jointly announced ten common principles for using artificial intelligence in new drug development, beginning to align standards on both sides of the Atlantic, which resembles adding one more artificial intelligence version to the good clinical practice standards already familiar to the pharmaceutical industry, the so-called GCP tradition.

Many evaluate that there is still a long way to go within Korea. Medical artificial intelligence experts point out that even in 2026, there are insufficient cases of using digital twins in the domestic medical field, and that a regulatory system to support trust must be established first. In the United States and Europe, regulatory authorities and companies accumulated data over 5 and 10 years respectively before presenting qualification recognition and guidance drafts. This gap resembles a problem of trust accumulation that takes time, rather than a problem of budget or will. However, that time cannot simply be allowed to pass.

The likelihood is high that the gap can narrow in a way where hospitals and companies that have accumulated sufficient data and established verification procedures first gain the trust of regulatory authorities, and their experience becomes the review criteria for the next case. The hybrid model in which digital twins and synthetic control groups replace half of clinical trials has now

entered the official documents of regulatory authorities. However, it appears that more time is needed to fill in explainability, data equity, and assignment of responsibility one by one until it becomes the default for all new drug development.

Representative Hong Suji, who leads the Chuncheon startup Offrime, points to the fact that foreign digital twin companies' virtual patient models were mostly built from Caucasian data as a differentiating factor of the company's technology. Offrime's artificial intelligence-based human digital twin, OPTIVIS, separately trains on data from East Asian patient populations including Koreans, and in a rare case globally, creates virtual patients of East Asian descent. Offrime first unveiled this clinical trial digital twin platform at the Alzheimer's Association International Conference (AAIC) held in London, United Kingdom in July 2026.

Given that developing a single new drug currently takes an average of 2 billion dollars and 10 to 15 years, Offrime announced that using OPTIVIS can reduce that time and cost by an average of approximately 30 percent. Offrime decided to build a new research base in Chuncheon Enterprise Innovation Park and signed a business agreement. The concern that data built primarily on Caucasians may not fit East Asian patients well appears to be a problem that will follow equally in multiple countries including Korea as digital twins spread globally.

Results mimicked with real-world data and actual clinical trial results are not always the same. Jessica Franklin and colleagues at Brigham and Women's Hospital affiliated with Harvard Medical School attempted to recreate ten completed randomized controlled clinical trials using only real-world data. Cases that reached the same conclusion as regulatory judgment were only six out of ten, and cases where the hazard ratio estimate fell within the confidence interval of the original clinical trial were eight. This result was published in the American journal *Circulation* in 2021. The accuracy between the two results

varied greatly depending on how closely the clinical trial design was reproduced to the original research. Before widely using digital twins and synthetic control groups instead of randomized controlled clinical trials, it appears safer to identify one by one under what conditions results diverge.

The more a patient desperately needs a digital twin, the less data that patient has. Rare disease patients and minority race patients have participated less in clinical trials up until now, and so they are underrepresented in the data used to train current digital twins. The academic community is attempting to statistically adjust for small numbers of patients or fill the gaps by creating new data with artificial intelligence, but cannot completely fill the limitations of insufficient original data itself. The risk that patients marginalized in old clinical trials will be marginalized again within new technology appears unlikely to disappear on its own unless new data is accumulated.

4.3

Precision Oncology and Personalized Treatment

The WINTHER clinical trial, conducted by the international WIN Consortium including Sheba Medical Center in Israel and MD Anderson Cancer Center in the United States, was a large-scale multinational study that enrolled over three hundred patients with advanced cancer, extracting both tumor and normal tissues to simultaneously read gene sequence and gene expression information. Administering a targeted anticancer drug called a tyrosine kinase inhibitor to multiple myeloma and breast cancer patients, the research team computationally analyzed the gene activity of each individual single cell collected from the patients' cancer cells. The patients were divided into two methods to be assigned treatments, with one side using DNA sequencing that read two hundred thirty-six genes, and the other side using RNA analysis that compared the gene expression of tumor and normal tissues.

The old method of choosing a drug by targeting only one specific gene had already reached its limit. Personalized medicine, which bundles each patient's different genetic information and pathological findings together to find a treatment that fits only that person, was moving to the center of cancer treatment. In the WINTHER trial, a multi-omics matching algorithm that calculated gene, protein, and metabolic information all at once to find the optimal treatment took on the role of pairing the patient with a drug. Tumors do not stand still. Chasing after a tumor that changes its form moment by moment inside a person's body was the question that this clinical trial posed.

Cancer drug development stood on one belief for a long time. It was the belief that if cancer occurred because one specific gene broke down, making a drug that accurately blocks just that one broken part would cure the disease. This

belief gained strength for a while after several drugs targeting specific genetic mutations actually produced dramatic effects. However, as time passed, cases accumulated where even patients with the same genetic mutation had divided responses, and drugs that worked well at first lost their strength after a few months or years.

The problem lay not in the drug but in the very framework of looking at cancer. As the awareness spread that a tumor was closer to a continuously moving video rather than a single photograph, and that a much larger picture encompassing the patient's genetic information, lifestyle, and immune status had to be read together, personalized medicine moved to the center of cancer treatment strategies.

The reason why even the same breast cancer differs from person to person is easy to understand if you think of students in a class. Just as students sitting in the same grade and same class have different heights, personalities, and subjects they excel in, patients with different genetic mutations are mixed together even under the single name tag of breast cancer. Tumors are no different. Although it looks like a single lump on the outside, looking inside reveals a mixture of cancer cell groups, each with slightly different genetic mutations, and this is called tumor heterogeneity. Just as students by the window and students in the back develop different habits even within the same classroom, the speed at which each cell grows and the way it responds to drugs change within a tumor as well, depending on its location and surrounding environment.

If we ignore this difference and treat the entire tumor as just one mass, we cannot explain why a drug that seemed to work well at first loses its strength as time passes. Therefore, even for two people who received the same breast cancer diagnosis, it frequently happens in actual examination rooms that a treatment that works well for one side does not work at all for the other side.

The process by which cancer cells run away to avoid medicine is a miniature version of natural selection. When an anticancer drug is administered, most cancer cells die, but a small number of cells possessing a genetic mutation that can coincidentally withstand the drug survive without dying. The surviving cells divide and multiply again, eventually filling the entire tumor with that mutation. Resistance is born this way. It is the moment when a drug that worked well at first no longer works. When one specific cancer-causing mutation appears, as its aftermath, the signal transduction pathway inside the cell is reorganized, the transcription environment that reads genes changes, and even the metabolic flow that creates energy is reassembled.

This kind of chain reaction cannot be put to rest by blocking just one molecule with a single drug. Recently, the research team is looking into the location and gene activity of each individual cell inside the tumor at the cellular level by mobilizing single-cell spatial immune profiling technology targeting Lynch syndrome patients. If the physical distance between stromal cells, immune cells, and cancer cells, and the non-linear interactions they exchange are analyzed with a deep learning model, the trajectory of which direction the tumor is evolving toward can be drawn in advance.

Among researchers, the perspective of seeing artificial intelligence as a single integrated engine that infers the evolution of the tumor itself is also gaining strength. This is an approach that entrusts artificial intelligence with the roles of aligning data obtained at different times into a single line, reconstructing the lineage of the cancer cell group, and even assessing the direction in which resistance will develop in the future.

All these predictions rely on the resolution of the data serving as raw materials. In the past, a single mass of tumor tissue was ground up to read the genetic signals of the numerous cells inside it lumped together as a single average value, but as single-cell sequencing and spatial transcriptome technologies spread, it has now become possible to record the gene activity of each

individual cell along with exactly where that cell was located within the tissue. It is akin to checking the names and grades of individual students after only looking at the class average score. It is only when data that has become this dense accumulates that artificial intelligence can also catch the minute differences of the cell group inside the tumor without missing them.

The complexity of the tumor grasped in this way leads directly to predicting drug responses for each patient. Here, multi-omics means looking at several layers inside the cell all at once, up to the genome which is the gene sequence, the transcriptome which shows how actively the gene is being read, and the proteins that are actually made. CDRscan, a deep learning-based model, takes a neural network trained by pairing anticancer drug response data confirmed in various cancer cell lines with the genomic mutation information of those cell lines as a foundation, to transfer the knowledge accumulated in the cell experiment stage to the prediction for the individual patient.

This is a method of calculating in advance with high accuracy how much a specific anticancer drug will actually be effective for a given patient, based only on cancer genomic signals. The PERCEPTION precision oncology pipeline, by learning both large-scale expression data obtained from large-scale cell line drug screening and single-cell expression data together, succeeded in predicting the response to targeted therapies and the occurrence of drug resistance in actual clinical trials targeting patients with multiple myeloma, breast cancer, and lung cancer. Both models essentially estimate treatment outcomes in a computer before placing a patient's tissue in the laboratory.

Recently, transfer learning methods in which learning is done first across multiple cell lines and that knowledge is then transferred to and applied in actual patient data are also being used to clarify the molecular causes of drug resistance. The work of integrating multi-omics data to pre-select patient populations that will respond well to immune checkpoint inhibitors is also carried out according to the same principle, greatly reducing the past

trial-and-error of confirming that a drug is ineffective only after administering the wrong medication and wasting several months.

Efforts to overcome drug resistance have advanced to the point of identifying the resistance pathways themselves. When artificial intelligence analyzed multi-omics clinical data from actual patients, it revealed that resistance to CDK4/6 inhibitors splits into pathways dependent on estrogen receptors and pathways independent of them. Because CDK4/6 inhibitors are a class of targeted therapies widely used for patients with hormone receptor-positive breast cancer, the achievement of revealing this resistance pathway has a significant impact on actual clinical practice.

VISTA is a protein that prevents immune cells from recognizing cancer cells as enemies, and STING is a signaling pathway that functions like an alarm device, detecting foreign substances that invade the body and turning on immune responses. Based on this discovery, researchers designed combination therapies that block VISTA or stimulate STING, filling gaps in drug resistance that could not be blocked by a single drug alone. This type of combination therapy targeting multiple targets at the same time is easy to understand when compared to cooking. Like a chef who does not try to create flavor with just one ingredient but calculates and adds together the combination ratio of multiple ingredients, artificial intelligence calculates both the combination and dosage ratio of drugs.

Deep learning techniques that predict synergistic effects between chemical compounds are also used for the same purpose, and because cancer cells attempt to survive by redirecting to another pathway when one pathway is blocked, a strategy is established from the beginning to target multiple targets simultaneously and block those alternative routes themselves.

There is also research showing how outcomes differ when increasing targets from one to multiple. The I-PREDICT clinical trial led by researchers at the University of California, San Diego campus used as a criterion how large a

proportion of multiple genetic mutations found in patients' tumors were targeted together by drugs, and showed results indicating that the higher the degree of match, the longer both progression-free survival and overall survival became. Targeting only a single mutation allows the tumor to quickly redirect to another pathway, but pressing multiple targets simultaneously reduces the escape routes proportionally.

As clinical decision support systems gather together and examine the features extracted by pathology foundation models from tissue images and genomic analysis results at once, the speed and accuracy with which medical professionals design multiple target combinations and verify their basis are increasing noticeably. Based on these results, researchers proposed that a design considering multiple targets together should be included in standard clinical procedures from the stage of initially conducting cancer genomic testing.

The way pathologists work - sitting in front of a microscope and reading tissue slides one by one with their eyes - is also changing. The term foundation model can be compared to the foundation work of a building. This is because it is a method of first laying broad foundational knowledge from millions of tissue pathology images and various types of clinical data, and then training specific tasks such as gastric cancer diagnosis or breast cancer prognosis prediction one by one on top of that foundation. Large-scale pathology foundation models created this way have begun to be actively used in cancer diagnosis and prognosis prediction.

These models convert the subtle shape of cells and the spatial arrangement of lined-up cells into numerical data called pathomics, and by combining this with genomic information or clinical records, they first simulate in a computer how patients with advanced gastric cancer will respond to preoperative chemotherapy or immunotherapy. Aiforia from Finland is artificial intelligence software that diagnoses gastric cancer, breast cancer, and prostate cancer,

and has successively received European in vitro diagnostic device certification (CE-IVD). Aiosyn, which was spun off from Radboud University Medical Center in the Netherlands, is also software that counts mitoses of breast cancer cells and received the same certification, reducing the time to read one slide by up to sixty percent.

Domestically, Lunit Scope, the pathology analysis platform developed by Lunit, is partnering with fifteen out of the top twenty pharmaceutical companies in the world and participating in clinical trials and biomarker development. In 2025, it formed a partnership in the digital pathology field with Labcorp, a U.S. diagnostic company, and presented joint research results at international conferences. In the same year, the European Society of Medical Oncology publicly released three studies that used Lunit Scope to predict immunotherapy treatment response together, providing evidence that pathology image analysis results are linked to patients' actual survival periods.

In the field of neuro-oncology, the AI-RANO guideline for evaluating treatment response in brain tumor patients in a standardized manner has been developed through an international research consortium, providing evidence that artificial intelligence-based image interpretation can be trusted in clinical trials.

Tempus, a U.S. precision medicine company, is a case that illustrates well the scale of this trend. This Nasdaq-listed company has accumulated over five hundred petabytes of data containing molecular information, imaging, and clinical records together, and within it are medical histories of over 45 million patients, genetic sequence information from 1.5 million people, and over four hundred thousand cancer patient records with complete genomic, transcriptomic, imaging, and clinical information. The Next platform put forth by Tempus is not an analysis that looks back at past data later, but proposes real-time clinical intelligence that provides doctors with the information they need at the very moment clinical care is being provided.

A study presented at the American Society of Clinical Oncology in 2026 released results showing that after introducing this platform to 662 patients across six American medical institutions, the biomarker testing rate for patients with early-stage non-small cell lung cancer clearly increased. In July of the same year, it announced the acquisition of Personalis, a company specializing in minimal residual disease testing, stepping into the work of tying together tissue-based precision oncology and blood-based residual disease tracking within a single company.

Tempus also started a national-level program called the PRECISION Challenge, which opens some of the research results of the foundation model it built to external researchers, widening its steps toward a direction of sharing data and models that a single company used to monopolize with academia.

The results produced by vast multi-omics data and predictive algorithms ultimately gather into one at the moment the doctor in charge issues a prescription. IBM Watson for Oncology, which learned the National Comprehensive Cancer Network (NCCN) guidelines by combining the knowledge of Memorial Sloan Kettering Cancer Center (MSKCC) and natural language processing technology, is a representative cognitive computing system that extracts patient information from electronic medical records and then presents treatment options to doctors in three stages: strongly recommended, for consideration, and not recommended. A framework called KM2PS reduces side effects by respectively checking the interactions that can occur between drugs and genes, drugs and drugs, drugs and health functional foods, and drugs and lifestyle habits, as well as liver function and kidney function, before prescribing anticancer drugs.

Artificial intelligence that determines clinical trial suitability, like OncoSOLVER, has also developed into a virtual platform that finds clinical trials fitting the patient in real time and supports the discussions of molecular tumor boards where doctors from various medical departments gather. The United States

Food and Drug Administration (FDA) operates a virtual tumor board through a decision support system called xDECIDE, where humans and artificial intelligence review cases together. This means that an attendee called artificial intelligence has been added to the existing tumor boards where multiple doctors used to gather.

Adding to this, the FDA is also developing a tool that automatically connects mutations found through next-generation sequencing (NGS) and drugs undergoing clinical trials targeting those mutations across electronic medical records, cancer knowledge bases, and clinical trial registration information. In 2026, research on an orchestration framework also appeared to support decision-making across oncology by tying multiple artificial intelligence models under a single chain of command, advancing to a stage of coordinating the judgments that different models produce over a single patient.

In one regional cancer care network, results also came out that the rate at which treatment decisions proposed by artificial intelligence matched the tumor board recommendations of a large cancer center increased from 60 percent before introduction to 94 percent after introduction.

Determining the treatment intensity also no longer depends on a single standard dosage table. A platform called CURATE.AI recalculated the drug dosage required for each solid tumor patient over time, proving the validity of an adaptive administration method that adjusts the dosage little by little while watching the patient's response in actual clinical practice. The idea is that even if the weight yesterday and today are the same, if the effects and side effects of the drug taken yesterday appeared differently, the dosage to be taken today must also be different. This method has a large use in pediatric cancer treatments or immunosuppressant adjustments after transplantation, where the gap between effect and toxicity is narrow and a single error leads to a large result.

In the field of nanomedicine, a virtual tumor model combining artificial intelligence with a model encompassing biological mechanisms at various levels numerically calculates the enhanced permeability and retention (EPR) effect, where nanoparticles pass through the blood vessel wall and accumulate in tumor tissues, to simulate in advance how the effect will differ for each patient when drugs are loaded and sent on nanoparticles. As this calculation becomes more sophisticated, grounds accumulate to determine the design and administration route of the necessary nanoparticles differently for each person, even for the same drug.

Finding cancer cells that might remain somewhere in the body even after removing visible tumors through surgery has long been a homework of precision oncology. As cancer cells die, they flow their DNA fragments into the blood, which is called circulating tumor DNA (ctDNA). A single drop of blood is the clue. Even at a point when nothing is yet visible in imaging tests, tracking the ctDNA concentration in the blood can gauge whether there is a relapse, and the traces of cancer left in the body after surgery like this are called minimal residual disease (MRD).

An artificial intelligence system called MRDsteer, reported to academia entering 2026, monitors the reliability of mutation reads in real time during the ctDNA analysis process, and runs an autonomous loop that selects only high-risk genomic sections to read again if the reliability drops below a certain standard. Verification data confirmed that this method is more stable than existing methods in catching even mutations of extremely low concentrations. The GEMINI Breast clinical trial, which repeatedly tracks ctDNA targeting breast cancer patients, has also been recruiting patients since October 2025, and various studies encompassing gastrointestinal cancer, lung cancer, and breast cancer are repeatedly confirming the results that patients who receive an MRD positive judgment after surgery or treatment have a higher risk of relapse and a shorter survival period.

Based on this information, medical staff can make decisions in a direction of increasing the anticancer treatment period for patients with a high risk of relapse, and lessening unnecessary treatments for patients with a low risk.

As cancer cells grow, they reveal altered protein fragments on their surface that were not in the original body, which are called neoantigens. Because neoantigens have completely different shapes for each person as much as the tumor mutations are different for each patient, they become markers present only in that person's tumor, much like each person's face. Artificial intelligence analyzes gene sequence data to select out those among these markers that are easy for immune cells in the body to recognize, and designs a vaccine just for that single patient based on this. When autogene cevumeran, jointly developed by BioNTech and Genentech, administered such a personalized vaccine to patients who underwent pancreatic cancer surgery, patients whose T cells were created in response to the vaccine had a significantly long period without relapse, whereas patients who did not respond relapsed in a median of 13.4 months.

In a follow-up study three years later, we also confirmed that the T cells of responding patients maintained a long lifespan averaging 7.7 years, and we are now broadening the scope with a Phase 2 clinical trial of 260 patients to verify the method of administering chemotherapy and immunotherapy together. mRNA-4157(V940), which Moderna and Merck are developing for melanoma patients, also reduced the risk of recurrence or death by approximately 49 percent when administered together with pembrolizumab, an immune checkpoint inhibitor. After announcing these clinical results, the companies entered Phase 3 trials, and it appears the decision on first regulatory approval will be made between late 2026 and 2027.

After confirming the tumor genetic sequence, it typically takes two to four weeks to actually produce this personalized vaccine. The cost of a personalized vaccine per person is estimated at between \$100,000 and

\$300,000. If used at the post-surgical adjuvant therapy stage for high-risk recurrence patients, the cost can be justified by the therapeutic benefit, according to analyses. Clinical trials combining immune checkpoint inhibitors and personalized vaccines have surpassed 150 cases, primarily in melanoma. BioNTech's BNT122, another personalized vaccine under development, is also advancing to Phase 3 trials in combination with immune checkpoint inhibitors for various solid tumors including colorectal cancer. The competition in personalized vaccines is not limited to a single company or a single cancer type.

When genomic and transcriptomic data, pathology images and radiological images, clinical records and circulating tumor DNA in blood, and lifestyle data - data of different forms - converge in one place, artificial intelligence synthesizes them to divide patients into detailed subtypes. That outcome in turn leads to personalized medicine and decisions on the administration route appropriate for each patient. Now it is no longer unusual for the department of pathology, the department of oncology, and the genomics analysis team to gather in one place to determine the treatment approach for a single patient. Domestically as well, this trend is no longer confined to the laboratory. The National Cancer Center is participating in international collaborative clinical research and accumulating precision oncology data, while major university hospitals including Seoul National University Hospital are conducting clinical trials combining chimeric antigen receptor T cell (CAR-T) therapy with liquid biopsy.

Four trends - targeted therapy, mRNA-based personalized vaccines, cell therapy, and liquid biopsy - have begun running simultaneously within a single hospital as of 2026, and the focus of treatment is shifting from grouping patients by a single disease name to using each individual patient as the standard. The way patients are referred to is also gradually changing. Instead of disease names like "lung cancer patient" or "breast cancer patient," language referring to patient groups with specific mutations - like

EGFR-positive lung cancer - and even to each patient's individual tumor signature is entering clinical practice.

However, no matter how sophisticated the formulas displayed are, the validation standards in clinical practice are far more stringent. The WINTHER clinical trial is an example demonstrating that reality. The trial succeeded in matching personalized medications to 35 percent of all patients based on omics data, but only 22.4 percent of patients met the pre-specified primary goal of having a progression-free survival 1.5 times or longer than previous treatment. However, when examining separately only those patients with good overall condition and high matching scores, the median overall survival was 25.8 months, showing a clear difference compared to 9.5 months in patients without such characteristics.

The gap does not disappear by itself. Deep neural network models that showed excellent performance within the training data often see their performance deteriorate when encountering new patient data from unfamiliar hospitals. This phenomenon is called the problem of illusory generalization or domain shift. Data leakage, where information about treatment outcomes that should only be known later accidentally gets mixed into the input values during model training and inflates performance, and shortcut learning, where the model learns only apparent correlations instead of true causal relationships, are cited as critical flaws that can lead to actual medical accidents.

Therefore, recent researchers have begun requiring external validation procedures - where models built in one hospital are re-validated on data from different hospitals, different populations, and different equipment environments - as a mandatory requirement for manuscript review.

Efforts to address the black box problem are also increasing concurrently. Models analyzing pathology images mark with heat maps which regions of the tissue images they used as the basis for their final determinations, allowing

pathology specialists to directly verify those regions visually. Models handling genomic data trace back numerically which genes and which interactions had significant impacts on the prediction results. While such explanation techniques are not perfect causal proof, they are serving as a bridge to gaining the trust of regulatory authorities by giving physicians room to judge the basis for AI recommendations for themselves rather than blindly following them.

The gap between regulation and technology carries the risk of escalating into legal disputes. Regarding whether synthetic patient cohort data designed by AI can substitute for control groups in masked randomized clinical trials, the FDA and EMA have been successively releasing guidelines in the direction of requiring that change management plans be predetermined based on risk level and disclosed transparently. The process of collecting single-cell multi-omics data and extensive clinical records and training from them conflicts with the European General Data Protection Regulation (GDPR) and the US Health Insurance Portability and Accountability Act (HIPAA) at multiple points, and this issue has yet to find a clear solution. If data is skewed toward specific races or patients from developed countries in the genomic data, the algorithm carries the risk of recommending incorrect diagnoses or inappropriate medications to marginalized populations, potentially widening healthcare disparities.

The black box problem, in which artificial intelligence cannot explain to a doctor the molecular-level reasons why it recommended a certain anticancer drug or why it warned of a specific side effect, still remains an obstacle delaying regulatory approval. When a patient is hurt or loses their life due to a mistake by an artificial intelligence clinical decision support system, no country yet has proper laws to decide whether the blame falls on the company that made the algorithm, the hospital that provided the data, or the doctor who wrote the prescription. The insurance reimbursement system is also failing to keep up with this change. This is because it is hard to find an answer for how to share and cover the costs of combination therapies and

personalized vaccines, which are different for every patient, using the old method of setting a price for a single standard treatment.

In the end, precision oncology in 2025 and 2026 can only move to the next step if it goes beyond the race to improve calculation speed or prediction accuracy, and also secures fairness in data, transparency in the reasons for its decisions, and a legal safety net in step with regulatory authorities. The fact that the person who makes the final decision is a human doctor has not changed amid all these technological advances.

In South Korea, precision oncology is no longer an unfamiliar word. Since June 2017, the K-MASTER project team led by Korea University has worked with 55 hospitals across the country to run the Match Master System, which connects the gene panel test results of cancer patients to ongoing clinical trials. Through this system, the doctor in charge checks whether there is a clinical trial that fits the patient and what targeted therapies can be considered.

Health insurance was also applied for the first time to next-generation sequencing (NGS)-based gene panel tests for 10 types of cancer in March 2017, and the scope widened to all cancer types in May 2019. However, the patient still has to pay half of the test cost themselves, and tests covered by insurance are limited to just two times: when first diagnosed, and when the cancer comes back or spreads to another place.

Finding a genetic mutation does not mean it immediately leads to treatment. There are cancer types, like lung cancer with a KRAS mutation, where the gene causing the cancer has been found, but a drug to stop that gene has not yet come out. Even when there is a drug, hurdles remain. Results showed that when Padcev and Keytruda, used for urothelial cancer, are given together, the cancer completely disappears in about 1 out of 3 patients, but this combination therapy does not yet receive health insurance coverage in South Korea, so the patient has to bear the entire cost of the medicine.

The rate at which new anticancer drugs enter South Korea within one year of their release stops at only 5 out of 100 cases, which does not even reach the Organization for Economic Cooperation and Development (OECD) average of 18 out of 100 cases. For targeted therapies used for lung cancer or breast cancer, it is not rare for a month's supply to cost more than several million won. The technology to find targets has become faster, but getting the right drug for that target and handling the price still remains a fight on a different level.

This gap also connects to the size of the hospital. Among patients newly diagnosed with cancer in 2023, less than half lived in the capital region, but nearly 8 out of 10 cancer patients who actually received treatment were at hospitals in the capital region. The number of cancer patients who live in regional areas but travel to the so-called Big 5 large hospitals in Seoul also increased by 11.8 percent in 2024 compared to 2022. The newest genomic tests and artificial intelligence-based analyses are usually set up first in a few large hospitals in Seoul, and it takes time for that equipment and workforce to spread evenly to regional hospitals.

The benefits provided by precision oncology are very likely to go first to patients who have the means to travel to those hospitals. Along with the work of raising the accuracy of the technology, widening the reach of that technology also appears to be a remaining task for the future.

PART 5

Regulatory Science, Data Ethics, and the Future Convergence Ecosystem

5.1

Regulatory Frameworks and Quality Governance for AI Drug Discovery

On January 14, 2026, a short document was released almost simultaneously at the Food and Drug Administration headquarters in Silver Spring, Maryland, United States, and the European Medicines Agency offices. Its name was "Guiding Principles on Good Artificial Intelligence Practice in Drug Development," and its contents consisted entirely of ten principles. Words like human-centered design, risk-based approach, data governance, and lifecycle management were densely packed into the document of just a few pages. This scene, where two giant regulatory agencies across the Pacific Ocean joined hands and spoke with one voice, was the moment they officially, albeit belatedly, acknowledged the fact that the center of gravity in new drug development is shifting from test tubes and animal cages to computer servers.

Eight months prior to this, in May 2025, the Food and Drug Administration announced that it would introduce Elsa, a generative artificial intelligence it developed internally, and by the end of June that year, completed its full deployment so that all reviewers could use it. As an internal report became known that Elsa finished a document comparison task in six minutes that used to take three days, the rumor that the speed of new drug review itself was changing was soon confirmed as fact. There are certain safety measures in the way Elsa is actually operated. The Food and Drug Administration confined Elsa within a government-exclusive cloud environment and designed it so that Elsa would not end up learning the sensitive review data submitted by pharmaceutical companies.

This essentially allows reviewers to access internal documents, while preventing their contents from flowing into Elsa's training data and affecting

the reviews of other companies. It means that the regulatory authorities themselves are applying the same standards to the artificial intelligence they use.

Regulatory agencies themselves call this change regulatory science. It means the task of redesigning the scientific basis itself for judging whether a new drug is safe and effective. Past regulatory science placed weight on statistically verifying the numbers that came out of clinical trials. Today's regulatory science must also embrace the task of verifying how reliable the artificial intelligence model that produced those numbers is. Chemistry and medical knowledge alone are no longer enough for reviewers, and they now need new eyes that can also read what data the model was trained on and what assumptions it operates on.

The picture running through all these movements is that regulations are changing from chasing after artificial intelligence to walking side by side with artificial intelligence. Just ten years ago, regulatory authorities made rules belatedly only after new technologies first came out in the market. Now, the Food and Drug Administration and the European Medicines Agency have changed direction to exchange opinions with the industry in advance over draft guidelines and to keep consultation windows open from the early stages of development. The role of the regulatory agency is essentially changing from an examiner demanding correct answers to a predetermined set of questions, to a partner solving problems together.

In December 2022, while passing the Food and Drug Administration Modernization Act 2.0, the United States Congress removed the mandatory animal testing clause that had been nailed down as a prerequisite for new drug approval applications for decades. Instead, the results of in silico simulations that calculate the movements of compounds in advance inside a computer, organoids which are microscopic organ models made of human cells, and organ-on-a-chip experiments that transfer the functions of human organs onto

semiconductor chips have stepped in as official evidence of safety and efficacy. The path to previewing with a computer how a new drug candidate will react inside the human body without going through mice and monkeys has essentially been opened by law.

This clause is closer to recognizing alternatives rather than immediately banning animal testing. Pharmaceutical companies can still choose animal testing, but the weight carried by the fact that the door has opened for the first time to stand on the review stage with only in silico data is different. How quickly they digest this response strategy is expected to determine the clinical development speed of each pharmaceutical company over the next few years.

The Food and Drug Administration had already prepared two mechanisms in 2021 when it released an action plan addressing artificial intelligence and machine learning software classified as Software as a Medical Device (SaMD). One is the ten principles called Good Machine Learning Practice (GMLP). It is a kind of recipe that dictates how to collect and share data, how to design and test the model, how humans and artificial intelligence will divide roles, and how to continuously monitor performance even after market launch. Among the ten, the principles that the data used for training and the data used for testing must be kept separate, that the data should evenly represent the patient population who will actually use the drug, and that the performance of humans and artificial intelligence working together, not artificial intelligence alone, should be examined are frequently cited.

The International Medical Device Regulators Forum (IMDRF) finalized and announced these principles as an international standard document in January 2025, and the regulatory phrases of the United States, Canada, and the United Kingdom essentially came together as one. The other is the Predetermined Change Control Plan (PCCP). Because artificial intelligence is software that continues to learn and change by encountering new data even after it is released, making it get approved again from the beginning every time it

changes would stop the entire industry. The Food and Drug Administration opened a path for manufacturers to predetermine the scope of possible future changes and verification methods to get approved all at once, and finalized final guidelines for medical device software containing artificial intelligence in succession between late 2024 and 2025.

It is similar to the method of agreeing in advance in the terms and conditions to what extent future software updates will be installed automatically when buying a mobile phone.

On January 6, 2025, the Food and Drug Administration released a draft guidance titled "Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drugs and Biological Products." The backbone of this guidance is a risk-based reliability evaluation framework made up of seven steps. Pharmaceutical companies must first reveal in writing what questions they are asking the artificial intelligence model, define the scope of what role the model takes at which stage of new drug development, and grade the size of the risk the patient will face when the model's judgment is wrong.

If they judge the risk is high, they gather thick evidence. They set up a validation plan and run it, then finally check if the result fits the original goal. If you explain this to an elementary school student, it looks like this. A teacher doesn't give a hundred points right away just because a classmate writes the right answer on a test. The teacher also looks at how they solved it. They check if the student just guessed, and if the student can get the right answer the exact same way on the next test.

Artificial intelligence that develops new drugs is no different. When artificial intelligence predicts a specific compound is harmful to the liver, the fact that the prediction is correct isn't enough. A person needs to be able to read the steps to see why it made that choice. They must be able to check if the same logic works consistently on other compounds. It's the same logic in a courtroom. Even if a witness tells the truth, a judge won't accept the statement

as evidence unless they look at how the statement came about and if it's credible.

The situation regulatory agencies fear is when artificial intelligence gets the right answer by chance, but no one can explain why. You can't repeat a judgment you can't explain. No one can promise what will happen when you apply an unrepeatable judgment to the next patient.

The documents the reliability evaluation system asks for have a few common items. These include source records showing where the data to train and validate the model came from. They include an uncertainty management plan showing in numbers how far the model's prediction is trustworthy, and when it starts to waver. They also include a life cycle monitoring plan to regularly check if performance drops after the model goes to market. This is a setup to filter out cases where people exaggerate their applications. They try to make a model that works well on just a few compounds look like it works on all compounds.

The attitude of changing the required level based on the risk appears in all the guidelines of the US, Europe, and Korea. Early-stage search artificial intelligence picks out promising candidates from thousands of compound candidates. It can pass with relatively light validation because a person checks the final judgment again. On the other hand, some artificial intelligence directly suggests the dose to give a patient, or it groups clinical trial participants into risk groups. This kind of artificial intelligence requires much thicker evidence and much more frequent revalidation because its judgment directly affects a human body.

Even if they share the name of artificial intelligence, the actual weight of regulation they carry changes greatly depending on how far their judgment reaches.

The same guidelines also cover introducing a system called the Model Master File (MMF). Similar artificial intelligence models are often used repeatedly in multiple new drug applications, just like locally acting medicines. If regulators make them submit new model validation data from scratch for every application, the burden of making and reviewing the same documents piles up for both pharmaceutical companies and reviewers. The Model Master File is a setup that separately registers a resume for an artificial intelligence model once it gets proper validation. In later applications, they can pull up that resume and use it as a reference document.

It looks like how you get a university diploma once. When you apply to multiple companies, you just submit a copy of your diploma instead of taking a test again every time. The industry and regulatory authorities are still discussing the detailed operation methods. But they point to it as a core setup to cut the inefficiency of overlapping model reviews.

If you look at the early stages of new drug development, artificial intelligence is changing the clinical trial design itself. In the past, researchers relied on experience and intuition to decide how many patients to gather and how to divide the doses. Now, a calculation model learns past clinical data and suggests a highly efficient trial design up front. People call this approach Model-Informed Drug Development. You can reach statistically reliable conclusions without blindly adding more patients. This advantage makes it very useful in areas where gathering patients is hard, like rare diseases.

It also cuts down the time and money it takes to find the optimal dose for a candidate substance. But regulatory authorities don't accept the design the model suggests right away. They check again within the reliability evaluation system procedure mentioned earlier. They look at how well the model's statistical assumptions match the real patient group's traits. They also check if the result's reliability dropped because they shrank the sample.

The EU AI Act was enacted in 2024 and starts taking effect step by step from August 2025. It classified artificial intelligence systems that predict the risk of clinical trial participants or control the quality of the medicine manufacturing process as high-risk artificial intelligence. Once classified as high-risk, a product must pass a separate conformity assessment before entering the European market. After launch, it must keep carrying massive regulatory compliance duties involving bias, transparency, and cybersecurity. The conformity assessment often doesn't stop at a self-assessment done by the company itself.

High-risk artificial intelligence medical products must go through a separate review by an independent review body the European Union designates, a so-called Notified Body. Without a certificate from this body, you can't put a product on the European market. But in May 2026, the European Union adjusted this schedule with a political agreement they call the Digital Omnibus. They pushed back the start date for high-risk artificial intelligence systems that count as medical devices by one year. The date moved from the originally planned August 2027 to August 2028. This move leaves the law's required goals the same, but gives companies more time to get ready.

In the meantime, medical devices that include artificial intelligence first get certification under the existing Medical Device Regulation and In Vitro Diagnostic Medical Device Regulation. The high-risk duties of the Artificial Intelligence Act then stack on top after the grace period ends. It operates as a double structure where two laws, not one, govern the same product one after another, and sometimes at the same time. Because of this, legal teams and regulatory affairs teams at pharmaceutical companies have to lay both schedules side by side and build a response strategy.

The European Medicines Agency first released a draft reflection document in July 2023 to gather comments, and then in September 2024 adopted a final reflection document addressing the use of artificial intelligence across the

entire lifecycle of pharmaceuticals. This document specifically warned about the phenomenon of hallucination, where large language models fabricate facts that appear plausible but do not actually exist when writing regulatory documents or summarizing clinical data.

The point was that the smoother a sentence appears to human readers, the greater the risk of accepting it as true without verification. Therefore, this document made clear that regardless of any artificial intelligence's judgment, final responsibility must always rest with humans. This means that even if artificial intelligence produces a highly sophisticated draft, it cannot be established as a regulatory document without a person to review and sign off on that draft.

No matter how tightly woven a regulatory framework is, it serves no purpose if the data supporting it is skewed in one direction. The genomic information and clinical records used to train artificial intelligence models are often concentrated in a few Western countries, and within those, among specific races and specific income groups. Applying models trained on such data directly to patients of different races or different genetic backgrounds increases the risk of errors such as missing unexpected side effects or recommending incorrect dosages.

Federated learning technology, which keeps original data on hospital servers and only exchanges the learned results, is attracting attention as a method for using data from multiple countries together without conflicting with laws that heavily protect patient information such as Europe's General Data Protection Regulation or the United States' Health Insurance Portability and Accountability Act. However, since each country has different computer systems and computational costs are substantial, more time appears to be needed before this technology becomes standard in actual regulatory documents.

The methods for validating traditional software and artificial intelligence are fundamentally different. With calculator programs, the same numbers always produce the same result, so once confirmed to be correct, there is no great need to test it again. However, because artificial intelligence is software that adjusts its weights slightly each time it encounters new data, there is no guarantee that a judgment that was correct yesterday will remain correct today.

Therefore, regulatory authorities treat artificial intelligence not as a test to pass once and be done, but as a living surveillance subject to be repeatedly verified as long as the product remains on the market. It resembles the structure of automobiles, which must undergo annual regular inspections rather than just one inspection when first purchased.

When an artificial intelligence model consisting of tens of millions of parameters identifies a specific compound as a toxic substance, even if the prediction fits perfectly, it cannot clear the threshold of regulatory approval if humans cannot explain the causal relationship behind that judgment. The work of technically explaining why accuracy alone is insufficient is the field known as validation and verification, or V&V. The industry frequently uses together a technique called Shapley Additive exPlanations (SHAP) and a technique called Local Interpretable Model-agnostic Explanations (LIME).

Both techniques take a single decision made by artificial intelligence and calculate in reverse how much each input value influenced that decision, presenting the results as numbers and images that humans can understand. While we cannot look inside the black box, this makes it possible to determine how much weight the black box placed on which inputs. In addition to this, the V&V 40 principles created by the American Society of Mechanical Engineers specifically for the medical field and the STARE-HI checklist, which is an evaluation and reporting guideline long used in the medical informatics field, are deployed together.

The core work is to cross-validate how well the model's predictions match the phenomena actually occurring within the human body, and to record the validation process itself in enough detail that other researchers can follow it exactly. It is the same principle as making grading rubrics public along with answer keys when grading test answers, so that anyone can re-grade by the same standards.

As artificial intelligence and digital twins have entered pharmaceutical manufacturing processes, Quality by Design (QbD), which is the principle of embedding quality from the design stage, and Process Analytical Technology (PAT), which monitors processes in real time, have begun to work together with artificial intelligence. It is a method in which sensors and artificial intelligence detect in real time the moment temperature fluctuates slightly or impurities enter, and automatically adjust control values. At this point, what regulatory authorities first verify is evidence that the data has not been tampered with. GAMP 5 2nd edition, revised by the International Society for Pharmaceutical Engineering in 2022, has carefully embedded the ALCOA++ principles, which refer to data accuracy, consistency, and traceability, into software validation practice.

Contained within this is the requirement that one must be able to identify who created the data and when, be able to read it, have it recorded at the exact moment the actual work occurred, have it preserved in its original form, and have it remain accurate and complete. One must keep an audit trail showing whether the data used for training and the data used for verification were physically separated, and whether there was no overfitting where the model answers correctly as if it had memorized the test questions in advance. It is no different from having to keep not only a chef's claim that they followed the recipe, but also receipts for the ingredients purchased and photographs of the cooking process in order to be recognized.

The habit of leaving detailed audit trails is how excellent machine learning practices are carried out not on paper but on the factory floor, and simultaneously leads to the practical competitiveness of pharmaceutical quality management departments.

When artificial intelligence independently searches through a knowledge graph and creates a molecular structure that has not existed before, the question of who should be granted the patent rights to that result is still not settled internationally. Because the patent system was originally designed on the premise of human creative activity, no country's patent office readily applies the same standard to inventions discovered by an algorithm alone with minimal human intervention. Different countries are reaching different judgments about whether to recognize inventions that cannot have a human name in the inventor field, and if recognized, whether to grant those rights to the company that created the program or to share them with the institutions that provided the data.

The question of who is responsible when an accident actually occurs also remains. Let us assume that a regulatory authority approved a new drug based on synthetic control group data or safety prediction results analyzed by artificial intelligence, but an unexpected patient death occurred after it went on the market. In this case, there is still no clear standard in any country to determine whether the software company that made the artificial intelligence, the hospital that provided the data, the doctor who issued the final prescription, or the regulatory authority that reviewed the documents and stamped the approval should bear the responsibility.

Just as debates are already taking place to divide responsibility among drivers, manufacturers, and software companies in autonomous driving accidents, artificial intelligence in new drug development will also likely face the same question in court soon.

The composition of people working in this field is also changing. In the past, chemists, doctors, and statisticians filled out the documents for new drug approval, but now data scientists, software verification experts, and regulatory lawyers sit around the same conference room. This is because assigning a risk grade to an artificial intelligence model and determining the locus of responsibility must ultimately be organized in legal language to complete the documents to be submitted to the regulatory authority. We are now in an era where not a single document can be properly completed unless those who know technology and those who know the law learn each other's languages.

After a new drug comes to the market, the role of artificial intelligence shifts to post-market safety surveillance, which is the area of pharmacovigilance. In the past, problems could only be known if doctors and patients voluntarily reported side effects, so rare and dangerous drug interactions were often discovered much later. Now, natural language processing technology scans electronic medical records, social media posts, and vast clinical literature in real time to catch subtle signals early that are easy for human eyes to miss. These materials that accumulate exactly as they are from hospital treatments and daily life are called real-world data.

The information visualization platform InfoViP and the Sentinel system built by the Food and Drug Administration are infrastructures that turn this real-world data into pictures so that reviewers can examine it at a glance. The Sentinel system was reorganized into version 3.0 with added natural language processing functions in January 2024, and it has conducted over 250 safety analyses so far. The Council for International Organizations of Medical Sciences (CIOMS) finalized and published the report of the 14th Working Group dealing with artificial intelligence and post-market surveillance on December 4, 2025. With seven principles as its framework - risk-based approach, human oversight, model robustness, transparency, data privacy, fairness, and governance - this report specifically organized how much artificial intelligence can be trusted in the practical work of precisely picking out rare side effects,

automatically filtering out duplicate reports, and automatically masking patients' personal information.

In this way, a circular structure is being completed in which safety information observed after marketing goes back and is reflected in the artificial intelligence training data of the early development stage.

As the Digital Medical Products Act took full effect in South Korea on January 24, 2025, a legal basis was established for the first time to group advanced digital medical products containing artificial intelligence, robots, and virtual reality technology into a separate category and handle them within a single framework from approval to post-market management. On the same day, the Ministry of Food and Drug Safety established and published the Generative Artificial Intelligence Medical Device Approval and Review Guidelines for the first time in the world. It was the result of forming a consultative body made up of experts from academia, the medical field, and industry to examine in advance what risk factors generative artificial intelligence medical devices hold from development through approval to post-market management.

This flow continued into 2026. The Ministry of Food and Drug Safety released a new civil petitioner's guidebook on June 30, 2026, containing the approval and review standards for digital medical devices based on large language models (LLM), and also overhauled the Change Management Plan Approval and Review Guidelines reflecting the characteristics of digital medical devices to which artificial intelligence technology is applied. This was a measure that paved the way so that if expected changes after approval are approved in advance, they can be directly reflected in the product without going through approval and review again each time, and its framework resembles the Predetermined Change Control Plan of the United States.

The Ministry of Food and Drug Safety also announced a plan to spend 34.7 billion won over the five years from 2026 to 2030 to train 1,100 regulatory science experts encompassing the fields of artificial intelligence new drug

development, advanced biotechnology, and digital medical devices. This budget will be divided and executed through ten implementing institutions covering degree and non-degree programs. It seems that the judgment underlying this investment is that if the reviewer does not understand the operating principle of the artificial intelligence model, no matter how elaborate the guidelines are made, they will lose their usefulness at the actual review desk. While the United States and Europe are establishing standards through laws and guidelines, South Korea is pouring its budget into raising the people who will actually handle those standards.

This trend also poses practical homework for South Korean pharmaceutical companies and bio-ventures. To apply for a new drug to the United States Food and Drug Administration, they must follow the reliability evaluation system of the United States, to enter Europe, they must jointly satisfy the high-risk classification of the Artificial Intelligence Act and the principles of the reflection paper, and domestically, they must follow the Digital Medical Products Act and the Ministry of Food and Drug Safety guidelines. The formats and terms of the documents required by the regulations of the three regions differ slightly, but the underlying question ultimately leads to one. If a domestic company builds up verification data according to international standards from the early stage of development, it can greatly reduce the trouble of repeatedly creating the same data when submitting applications simultaneously to multiple countries later.

The International Council for Harmonisation (ICH) finally adopted a new guideline, M15, dealing with the general principles of model-informed drug development on January 29, 2026. This is the work of creating a language that regulatory authorities in each country can commonly refer to when computational models, including artificial intelligence, are used for clinical trial design and dose determination. Although there is no independent International Council for Harmonisation guideline that deals with artificial intelligence separately yet, if you look at it together with ICH E6(R3), the

principle of clinical trial design that took effect from July 23, 2025, you can see that various guidelines are interlocking in one direction.

The United States Modernization Act and reliability assessment system, the European Union's artificial intelligence law and reflection documents, international organizations' best machine learning practices and post-market surveillance principles, and South Korea's Ministry of Food and Drug Safety's laws and personnel training plan each carry different document names, but the core requirement they demand converges into one. Beyond whether the answer produced by artificial intelligence is correct, the principle is that humans must be able to verify the process by which that answer is reached and take responsibility for it. From a lawyer's perspective, this is not a strange requirement. The long-standing courtroom practice of questioning the credibility of evidence and requiring that the basis for decisions be recorded is simply now being transferred into the data pipelines of laboratories, factories, and hospitals.

However, there are also differences from courtroom practice. The court looks at events that have already occurred and makes judgments after the fact, but new drug regulation is an area of prevention that must establish thorough verification procedures before accidents occur. The question of how well humans can understand and take responsibility for results produced by artificial intelligence will likely be asked repeatedly at each stage of new drug development in the future.

The movement to fill the gap left by the absence of standards for determining liability first appeared in Europe. In November 2024, the European Union revised and promulgated the Product Liability Directive, deciding to treat software and artificial intelligence systems as products just like tangible goods and to impose strict liability on them. Member States must transpose this directive into domestic law by December 9, 2026, and injured patients have gained a broader path to claim damages without having to separately prove

the developer's negligence. However, since the development risk defense - that a defect could not have been discovered with the level of technology at that time - also applies to software, a new path for artificial intelligence developers to escape liability has also opened.

South Korea's Product Liability Act does not yet explicitly include software in the scope of products, so even if an accident occurs with a new drug designed by artificial intelligence, holding the software itself liable remains in a state with substantial room for dispute in court. If domestic legislation fails to keep pace with Europe's speed, there is a strong possibility that Korean patients who experience the same accidents will find themselves bearing a heavier burden of proof.

The question of whether reviewers have the capacity to verify artificial intelligence receives an even colder answer from workforce statistics. On April 1, 2025, the United States Department of Health and Human Services laid off 3,500 Food and Drug Administration employees at once, and reports followed that the remaining reviewers had to take on work that their colleagues had been handling. This was less than three months after announcing the risk-based reliability assessment system. Apparently to fill that gap, the Food and Drug Administration deployed a new agentic artificial intelligence accessible to all staff on December 1, 2025, and this tool became able to share responsibilities including pre-market review, verification of review results, and post-market surveillance activities.

The approach of filling insufficient review personnel with artificial intelligence appears to contradict the principle that humans must verify the judgments made by artificial intelligence. South Korea's Ministry of Food and Drug Safety faces a similarly challenging situation. As of November 2025, the Ministry of Food and Drug Safety has 369 medical product review personnel, which differs by an order of magnitude when compared with the United States Food and Drug Administration's 9,000 review personnel, the European Medicines

Agency's 5,500, and Japan's Pharmaceuticals and Medical Devices Agency's 1,300. Although the Ministry of Food and Drug Safety is discussing with the Ministry of Interior and Safety ways to hire an additional 295 reviewers, preparations remain slow, to the point that observers have pointed out that the website announcing recruitment information has been inactive since 2021.

No matter how much budget is allocated to personnel development, if there are so few people sitting at the review table itself, it appears that the opportunity to examine each artificial intelligence model closely will differ greatly from country to country, and it appears that companies will also emerge that adopt the strategy of submitting applications first to countries with lax verification.

5.2

Data Ethics, Algorithmic Fairness, and the Federated Learning Ecosystem

In July 2024, more than ten pharmaceutical companies and university research teams gathered at the office of the Korea Pharmaceutical and Biotech Industry Association in Seoul. The project name was K-MELLODDY, a five-year national initiative running through December 2028, and the plan was for the Ministry of Health and Welfare and the Ministry of Science and ICT to invest 34.8 billion won to accelerate new drug development based on federated learning. The atmosphere in the conference room was peculiar. Name plates from each institution were placed on the table, but not a single shared screen appeared on the laptop displays to show each other.

Although competing pharmaceutical companies and researchers came together in one place, none of them wanted to hand over their company's chemical compound data to competitors. Yet they joined hands to train an artificial intelligence model together. Without exchanging even a single line of data. To understand why this seemingly contradictory promise actually works, we need to first look at the fundamental dilemma that new drug development artificial intelligence faces right now.

To achieve precision medicine, we need to gather the genomes and medical records of millions of people. A single hospital or pharmaceutical company cannot identify patterns in rare diseases with only the number of patients they have. However, the moment patient genetic information and medical records are gathered in one place, that database becomes a target for hacking and data leaks. The United States strictly prevents cross-border transfer and external disclosure of patient information through HIPAA (Health Insurance Portability and Accountability Act), and the European Union does the same

through GDPR (General Data Protection Regulation).

From the pharmaceutical company's perspective, the data needed to create a good artificial intelligence model lies beyond the hospital walls, but the law says not to cross those walls. If we cannot resolve this deadlock, artificial intelligence-based new drug development will remain a closed game where only a few large pharmaceutical companies grow using their own internally held data.

The technology that broke through this deadlock is federated learning. To explain it in a way that even an elementary school student can understand: ten classmates are studying for a test, and each of them promised never to show their notebook to anyone else. Instead, each evening after checking their solved problems against the correct answers, they tell each other only which problem-solving techniques were more effective at raising their scores. After ten days, all ten of them have not seen even a single word of their classmate's notebook, but their individual test scores have improved compared to when they studied alone.

Federated learning works the same way. Hospitals and pharmaceutical companies never send out the notebook containing patient data beyond their own server. Instead, after the artificial intelligence sees that notebook and gradually becomes smarter, it encrypts only the model weights, which are composed of numbers, and sends them to the central server. The central server combines the weights from various institutions into one and returns a more intelligent model to each institution. This procedure of averaging the weights from multiple institutions and combining them into one is called federated averaging.

This method of weight exchange does not guarantee safety in theory alone. It is because attacks are possible to reverse-engineer and estimate the original data from the weights themselves. Therefore, recent researchers are combining differential privacy techniques, which intentionally mix

fine-grained noise before sending weights to the central server, with secure multi-party computation techniques that prevent multiple institutions from seeing each other's input values while only matching computational results.

The PrivMol study published in 2026 combined these two techniques in a graph artificial intelligence model that predicts molecular properties, showing results that lower the risk of original data leaking without greatly reducing prediction accuracy. It is like locking a lock not with just one lock but with double and triple locks.

The first project to prove that federated learning actually works in new drug development sites was MELLODDY in Europe. Led for three years starting in 2019 with 18 million euros jointly provided by the European Innovative Medicines Initiative (IMI2) and the European Pharmaceutical Industry Association (EFPIA), this project involved ten global pharmaceutical companies and seven technology and academic institutions. The data they used together for learning included more than 2.6 billion experimental activity values, more than 21 million compounds, and more than 40,000 biological test results. No company showed its compound list to competitors.

Yet the joint model created through federated learning showed predictive performance improvements of an average of 5.5 percent, and up to 9.7 percent, compared to models each company created with only its own data. Even though they kept different ingredients in different refrigerators, they refined one recipe together and improved everyone's cooking. This improvement margin means that they could filter out that many more failed compounds without directly synthesizing them in the laboratory, which together reduces the time and cost of the early stages of new drug development.

The same idea was reborn in Korea under the name K-MELLODDY. This project, which the Korea Pharmaceutical and Biotech Industry Association launched in July 2024, is structured around three pillars: creating a federated learning

platform, gathering and supplying data from hospitals and pharmaceutical companies, and training artificial intelligence models with that data. The result the project team aims to produce is a prediction model called FAM (Federated ADMET Model). ADMET is a term that collectively refers to how well a chemical compound is absorbed when it enters the body, how it spreads, how it is broken down, and how much toxicity it causes.

In 2025, various institutions including Korea University, Soongsil University, Korea Advanced Institute of Science and Technology, and LG Chemical took on and began five new subtasks. From the perspective of mid-sized domestic pharmaceutical companies and bioventures, which often lack the capital to purchase data, this method of maintaining ownership while learning together is evaluated as attractive.

Beyond cooperation among pharmaceutical companies, federated learning that connects hospital to hospital is also growing rapidly. With rare diseases, the number of patients gathered in one country or one hospital is so small that there is a shortage of samples from which artificial intelligence can learn patterns. For example, in multinational hospital consortiums that handle tumor imaging or pathology photographs, hospitals from different countries exchange only the weights of the model that distinguishes tumors while keeping their own patients' images within their own servers.

While no hospital sends original images that reveal patient identities to other hospitals, the model reaches an accuracy it could not have achieved with just one hospital's data. This cooperation is possible because the original data does not physically cross borders.

The value that federated learning protects and the value that explainable artificial intelligence protects are like two weights placed on opposite pans of a scale. On one pan is the weight of security. Federated learning presses down on this pan by allowing only knowledge to travel, without moving the original data a single step from where it was. On the other pan is the weight of

reliability. XAI (eXplainable AI, explainable artificial intelligence), which explains to humans why the artificial intelligence made a certain judgment so they can understand it, presses down on this pan.

If either side becomes heavier, the scale will tilt. If the data is kept safe but no one can explain why the model reached that conclusion, neither doctors nor patients can trust and use the results. On the contrary, if the explanation is perfect but the data itself that produced the explanation is exposed to the risk of leaks, that artificial intelligence cannot cross the hospital threshold in the first place.

However, just because data is gathered safely does not guarantee that the data itself is fair. Many of the clinical trials and genome databases scattered around the world are skewed toward WEIRD populations, that is, the populations of Western, educated, industrialized, rich, and democratic nations. As of mid-2025, the proportion of populations with European ancestry among the people who participated in genome-wide association studies (GWAS) reached 90.53 percent. The share of populations with African ancestry has hardly increased over the past decade.

For more than 15 years since 2009, the proportion of research centered on European descent has only fallen by roughly 5 percentage points. If a nanoparticle drug delivery model or a tumor target prediction model trained solely on data from white patients in North America and Europe is applied directly to patients in Asia, Africa, or South America whose genetic and lifestyle backgrounds are completely different, the prediction performance drops sharply, and the risk of unexpected toxic side effects increases.

Multiple artificial intelligence diagnostic tools that detect skin cancer showed a sharp drop in diagnostic accuracy for patients with dark skin tones, because they were trained mostly on photos of patients with light skin tones. It is not just a problem with artificial intelligence software. The pulse oximeters widely used during the COVID-19 pandemic also showed errors that measured the

blood oxygen saturation of patients with dark skin tones higher than the actual levels, prompting the Food and Drug Administration to issue a separate safety warning. One artificial intelligence tool that diagnoses bacterial vaginosis had high accuracy for white women, but its accuracy dropped sharply for Asian women, and it produced a uniquely high number of false positives for Hispanic women, incorrectly determining that they had the disease when they did not.

A risk prediction algorithm used by a large insurance company in the United States systematically pushed Black patients out of its health care program. It used medical expenditures as a proxy indicator for health risks, because Black patients had lower medical accessibility and thus lower expenditures even when they had the same diseases. The algorithm perfectly copied discrimination based on race without ever looking directly at race. Technically, adversarial learning techniques that train models to purposefully counter biased patterns are also being used as a method to reduce bias.

There are also studies that reveal data equity in numbers. When a skin disease diagnostic model was analyzed using the HEAL framework, which measures the health equity of machine learning models, the equity score was an excellent 92.1 percent in subgroups based on gender, but the score dropped all the way to 0.0 percent in the group of older patients suffering from skin diseases that were not cancer. This means that within the same model, some patients were being diagnosed precisely, while some patients were effectively being neglected. Another trap that occurs when the balance of training data is broken is shortcut learning.

If you train a model with data where the sample of healthy people is 80 percent and the sample of sick people is 20 percent, the model achieves a plausible accuracy of 80 percent simply by unconditionally guessing that every input value is healthy. An optical illusion is created where the model appears excellent just by looking at the numbers, even though it actually fails to find a single disease. When caught in this trap, patients with less data, such as those

with rare diseases or people of color, entirely bear the risks of misdiagnosis and inappropriate prescriptions.

The Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have firmly embedded a diversity action plan into their clinical trial protocols, which requires intentionally including racial minorities, diverse age groups, and socioeconomically marginalized classes right from the stage of designing clinical trials. In August 2025, the Food and Drug Administration released guidelines establishing how to manage medical device software containing artificial intelligence from before to after it goes on the market. These guidelines require the representativeness of training data, bias management, cybersecurity, and post-market performance monitoring, while establishing a system called the Predetermined Change Control Plan (PCCP) that allows algorithms to be modified without receiving new authorization, as long as it is within a predetermined scope.

In 2026, separate guidelines regarding clinical decision support software were also released. The principle that we must not stop at simply increasing the amount of data, but must selectively gather high-quality data that actually represents the population composition, seems to be moving beyond the recommendations of a few pages of documents and becoming a practical standard in the review process for new drug approvals.

Attempts to fix data skewness at its roots are also proceeding at the same time. Until now, human genome research has relied on a method that takes the genome of one person from a specific population group as the standard reference sequence, and compares the genomes of all other people against that sequence. Because this single reference sequence was not designed to reflect the diversity of humanity in the first place, population groups with more genetic variations that deviate from the standard became more disadvantaged in diagnostics and treatment predictions.

Recently, a pangenome reference map that overlaps the genomes of multiple population groups has emerged as an alternative to compensate for this limitation. If samples from population groups that were marginalized in previous research are abundantly included in the pangenome map, artificial intelligence models can use more even data as their learning materials from the very beginning.

Even if the data bias is corrected, there is one more wall to overcome inside the deep learning model. It is the black box problem. A neural network entangled with tens of millions to billions of parameters instantly produces a conclusion that a certain compound is harmful to the liver or promising as an anticancer target. However, in many cases, even the researcher who designed it cannot explain what path led to that conclusion. It is like putting compound information into a box that cannot be opened and having only the answer pop out.

No one can see what calculations went back and forth inside the box. In the drug approval review, where a patient's life is at stake, such opacity is difficult to accept as it is. For a doctor to prescribe and regulatory authorities to stamp an approval, a single number called predictive probability is not enough, and they must be able to look into the judgment process behind it.

The attempt to open this black box is explainable artificial intelligence, or XAI. Techniques that retrace the artificial intelligence's conclusion, such as SHAP or LIME, visually depict with colors or numbers how much a certain chemical structure of a compound or a certain sequence of a protein contributed to predicting binding affinity or toxicity. Instead of just presenting a finished piece of furniture, it is similar to handing over an assembly manual written with how many screws were driven where. Only after seeing this manual can doctors and regulatory authorities verify whether the conclusion came from actual biological evidence, rather than an accidental statistical correlation.

However, a 2026 study evaluating XAI for antibiotic discovery pointed out the limitation that the standards for measuring the performance of these explanation techniques are not yet unified. While the tools that provide explanations are increasing, the yardsticks for measuring how reliable those explanations are have not yet been fully equipped.

For an artificial intelligence prediction model to take root in the actual clinical field, such an explanation technique alone is not enough. Procedures to verify the reliability of the computational model, such as the V&V 40 principles set by the American Society of Mechanical Engineers, and supervisory procedures where humans reconfirm the final judgment must be woven together from the design stage. Recently, reporting guidelines that require transparent documentation of the artificial intelligence model's performance, training environment, and clinical application scope, such as CONSORT-AI and DECIDE-AI, have also been increasing, mainly in the academic community.

Artificial intelligence medical devices that enter the market skipping those procedures are increasingly being caught up later because they cannot provide supporting data during audits or lawsuits. However, these guidelines are still closer to academic journal editorial policies and are not norms enforced by law. This gap between recommendation and obligation leads directly to the question of who to hold legally responsible when an actual medical accident occurs.

As artificial intelligence goes beyond diagnosis and prediction and reaches the level of designing new compounds on its own, a completely different legal issue has opened up. Even without a researcher directing it one by one, generative artificial intelligence learns data containing the relationship between structure and activity and protein folding information on its own to design a new molecular structure from scratch. This is the so-called de novo design. Compounds where artificial intelligence was involved in the design from start to finish have already entered the clinical stage in several cases,

such as the case where Insilico Medicine pushed a candidate for idiopathic pulmonary fibrosis treatment designed by its own artificial intelligence platform to phase 2 clinical trials, or DSP-1181, an obsessive-compulsive disorder treatment candidate designed by Exscientia that entered phase 1 clinical trials in 2020.

The problem is whose name should be written in the inventor box when filing a patent for a molecule born this way.

There is a case that met this question head-on. American inventor Stephen Thaler claimed that an artificial intelligence named DABUS made inventions on its own, and he filed patents naming DABUS, not a human, as the inventor in several countries, including the United States, Europe, the United Kingdom, and South Korea. In 2022, the United States Court of Appeals for the Federal Circuit in the Thaler v. Vidal case did not recognize DABUS's status as an inventor, pinning down that an inventor under United States patent law must be a natural person. On November 28, 2025, the United States Patent and Trademark Office issued new examination guidelines dealing with the inventorship of inventions involving artificial intelligence, reaffirming the principle that artificial intelligence is only a sophisticated tool like a laboratory instrument and that conception, which is the core of an invention, is an act that only humans can perform.

On March 4, 2026, even the Supreme Court of Japan dismissed Thaler's appeal, effectively ending the worldwide patent battle over DABUS that had dragged on for seven years. Looking back in time, the European Patent Office in 2021 and the Supreme Court of the United Kingdom in a December 2023 ruling already reached the same conclusion. Until now, the patent offices and courts of almost all countries, including the United States, Europe, the United Kingdom, and Japan, have not recognized DABUS as an inventor.

Looking through the eyes of a lawyer, the principles spoken by these precedents are clear. Even if artificial intelligence actually designed a

candidate substance, the name of the researcher who designed and handled the artificial intelligence and actually recognized the value of the compound must be written in the inventor box on the patent documents to protect the patent. The problem is how the companies that spent hundreds of billions of won to build an artificial intelligence drug discovery platform will translate this principle into practice. This is because when artificial intelligence pours out millions of candidate molecules overnight, they must prove one by one in every patent application document to which one and how creatively the researcher intervened.

If this proof fails, there is a risk that the painstakingly discovered compound will be left exposed to competitors without being protected by a patent. Conversely, if the researcher's contribution is exaggerated too much, it is like creating an excuse for the patent itself to be overturned and invalidated later. For this reason, some companies choose a strategy of protecting their compounds as trade secrets instead of patents from the beginning, but trade secrets have a limitation that the protective shield disappears as is if a competitor independently finds the same compound. The practice of striking this balance is highly likely to remain a core issue in patent litigation for years to come.

Conflicts over intellectual property do not end with the question of inventor status. As research increases in which artificial intelligence uses large language models to decode traditional medical texts and Indigenous communities' knowledge of medicinal plants to find new natural product drug candidates, the question has arisen together of what rights to recognize for the communities that have protected that knowledge for centuries. The Nagoya Protocol, adopted in 2010, is an international treaty that stipulates that those who wish to use genetic resources and associated traditional knowledge must first obtain the consent of the communities that hold that knowledge and fairly share the profits generated from that use.

So far, 142 countries have ratified this treaty, and over 130 countries each have laws related to benefit-sharing. However, now that artificial intelligence rapidly learns vast traditional knowledge literature and proposes new compounds, procedures for tracking which materials came from which communities and sharing benefits have not yet been properly established. The benefit-sharing agreement between South Africa's San and Khoi peoples and the rooibos tea industry is frequently cited as a case where royalties were actually paid to communities holding traditional knowledge.

Voices in academia are growing that when artificial intelligence reviews traditional knowledge literature to find new compound candidates, royalty payments and source attribution at this same level should be implemented. The incentive to convert traditional knowledge into privatized patents and reap enormous profits collides directly with the demand that communities that have protected that knowledge be given fair compensation.

Beyond the patent issue, an even sharper question awaits. It is the question: when a patient is actually harmed due to a decision made by artificial intelligence, who bears responsibility? Even when a clinical decision-support system is validated using explainable artificial intelligence and trained through federated learning, subtle biases hidden in the training data can cause it to incorrectly recommend a medication as safe when it is not safe for certain patient populations. For example, if the training data did not adequately capture side effects that appear only in patient populations with a specific genotype, the artificial intelligence could render a safety judgment without predicting that risk.

A solid legal framework that can distinguish whether responsibility lies with the information technology company that designed the neural network, the hospital that provided the data, the physician who prescribed based on that recommendation, or the regulatory authority that granted approval has not yet been fully established anywhere in the world.

The European Union prepared separate guidelines on artificial intelligence liability to fill this gap, but in February 2025 announced it would withdraw the draft guidelines, and in October of that year confirmed the withdrawal through publication in the Official Journal, changing course to address artificial intelligence-related accidents within the revised Product Liability Directive. The European Union's Artificial Intelligence Act itself took effect on August 1, 2024, and from August 2025, fines can be imposed on those creating general-purpose artificial intelligence models. The obligations for high-risk artificial intelligence systems including medical devices were postponed through the 2026 Omnibus amendment, with high-risk systems corresponding to Annex 3 applying from December 2, 2027, and Annex 1 systems embedded in products applying from August 2, 2028.

Companies handling high-risk artificial intelligence must establish a risk management plan, quality training data, transparent documentation, human oversight mechanisms, and continuous monitoring systems after deployment, and if any of these is lacking, there is room for it to be treated as a defect as described in the Product Liability Directive. It is a structure in which obligations are set, and if those obligations are violated, it becomes difficult to escape responsibility when an accident occurs.

The European Union's Artificial Intelligence Act regulates medical artificial intelligence with a double binding. Artificial intelligence systems involved in patient diagnosis or treatment decisions must obtain certification according to medical device regulations while also complying with the high-risk system obligations set out in the AI Act. If either of these two regulations is not met, market entry itself is blocked. This means that algorithm performance alone is not enough; the data that produced that performance and the design process from beginning to end must be documented in order to enter the market.

Among institutions participating in federated learning consortiums, the practice of pre-determining liability and insurance responsibility in

participation agreements is also increasing. This is a movement to distribute risk through contractual provisions before disputing liability in court when an accident occurs.

Korea's legal system is also moving in a similar direction. De-identified information refers to personal information in which items that can directly identify a specific person - such as name or resident registration number - are deleted or replaced with other values, making it difficult to identify who it is without additional information. Korea's Personal Information Protection Act stipulates that de-identified information held by different institutions cannot be directly compared by personnel, and that data can only be combined through a third-party institution called a designated data combination agency. Following a research and development roadmap running from 2022 to 2026, the government has been refining systems and technologies for combining de-identified information held by various institutions, and in 2026 will even hold a competition comparing cases of de-identified information linkage.

The Ministry of Health and Welfare announced a plan to connect medical data and artificial intelligence technology to drug development and medical settings throughout the 2026 research and development project, and the government decided to invest 400 billion won through 2030 in artificial intelligence-based drug development and advanced brain science research.

In Europe, the European Health Data Space (EHDS) regulation took effect on March 26, 2025, as a corresponding system. According to this regulation, the entire regulation will apply from March 26, 2027, and primary use for individual care and secondary use for different purposes such as research, policy, and artificial intelligence development will both be implemented in stages over two periods in 2029 and 2031. Patient summaries and electronic prescriptions and most research data will be opened first in 2029, while sensitive or substantial data such as medical images, test results, discharge records, and genomic information are postponed to 2031.

Each member state is established with a health data access body that issues permission for data use and manages responsibility. The European Data Protection Board and the European Data Protection Supervisor were concerned that this regulation risked weakening the individual rights established by the General Data Protection Regulation. As a result, the final provision was compromised by placing the legal basis for secondary use back on top of the General Data Protection Regulation. Within one regulatory body, the regulation protecting privacy and the regulation seeking to gather and use data are pulling tightly against each other.

The reason the researchers gathered in the K-MELLODDY conference room were able to improve their grades together without ever showing each other their notebooks was that they proved technology can protect two values - trust and security - at the same time. However, behind the door that technology opened, there remain unresolved questions. These are matters of determining which population group's data the model was trained on, whose name the patent for compounds designed by artificial intelligence bears, and which institution bears responsibility when an accident occurs.

From the perspective of patients and citizens, it is worth remembering that when this balance breaks down, those first exposed to risk are always people with little data and who find it difficult to raise their voices. This work does not end within laboratories but continues to move forward to courts, the National Assembly, and the negotiating tables of international organizations. It appears that including bias audit clauses and requirements to record contributions to inventions in contracts at this stage now will reduce the cost of determining responsibility when disputes arise later.

Just as the courts in the United States and Japan have settled on not recognizing artificial intelligence as an inventor, there is a high possibility that over the next several years, norms addressing data equity and assignment of responsibility will be built up country by country through case law and

guidelines. Until those norms become dense, the balance of trust and security that federated learning and explainable artificial intelligence are creating now appears to continue to be tested.

On December 31, 2025, the Ministry of Health and Welfare and the Personal Information Protection Commission revised the guidelines for the use of healthcare data and put them into effect. A new shared DRB system was created for small and medium-sized institutions that struggle to establish their own institutional health and medical information review committee, that is, a DRB, or to conduct research jointly with multiple hospitals. Standard procedures were also established to align the review procedures of the institutional bioethics committee and the DRB into one.

Standards were newly established to erase identifiers that reveal identity and codes related to surviving family members when using medical data from deceased patients in research. For domestic startups or university research teams that use artificial intelligence to identify new drug candidates, it appears that the review waiting time that was needed until collecting data from multiple hospitals and training a predictive model will be reduced by this revision.

However, all these procedures are conducted only under the eyes of the review committee, not from the perspective of the patients who provided the data. The domestic personal information consent system remains at the level of obtaining consent once when a hospital initially collects information, and there is no pathway for patients to later trace back which research or which company's artificial intelligence learning the pseudonymized information was used for. Moreover, the law has already permitted that pseudonymous information need not obtain additional consent from the information subject even when used differently from its original purpose, such as for statistical compilation or scientific research.

Some countries, like the UK National Health Service, have opt-out mechanisms that stop information use when information subjects indicate their refusal even at a later time, but Korean law has not yet reached that stage. That consent is not necessary does not necessarily mean that monitoring is unnecessary, but it is unclear who will be responsible for that monitoring. No matter how safe federated learning that does not move data from where it originally is, if patients have no way to verify where their information goes, that safety risks remaining incomplete.

There are gaps even in the belief that federated learning protects original data. In March 2025, the international academic journal *Journal of Cheminformatics* published results of a research team led by researcher Fabian Kruger attempting a membership inference attack against a neural network model that predicts the properties of new drug candidate substances. The method was to guess whether a specific compound was in the training data by simply asking questions from the outside and receiving answers without looking at the weights inside the model.

In any dataset and model structure, information leaked, and compounds belonging to minority groups with small sample sizes were more easily exposed. In drug development, compounds with small samples are often valuable candidates that competitors have not yet touched, so there is a risk that the very information that should be protected is leaking first. The Kruger research team publicly released the attack technique as code, allowing other researchers to check the risks of their own models.

5.3

The Future of the Pharmaceutical Industry: Self-Driving Labs

In October 2025, a laboratory in Seoul quietly opened its doors. In this space created by the AI New Drug Research Institute under the Korea Pharmaceutical and Bio-Pharma Manufacturers Association, a single robotic arm stands in front of a scale instead of researchers wearing white coats. The robotic arm twists open a glass bottle cap and accurately portions out 0.005 milligrams of a reagent, an amount corresponding to a fraction of a hundredth of the weight of a grain of salt. The actions of mixing liquids and weighing them again are repeated at intervals of a few seconds. After ten o'clock at night, the researchers leave work one by one, and the lights in the laboratory are turned off. The robot does not stop in that darkness.

People call this space a 'dark lab'. An association official explained that the goal of this laboratory is to maximize experimental throughput even while people are sleeping. For every reaction vessel handled by the robot, the experimental conditions and results are automatically recorded, and the researcher arriving at work the next day begins the day by checking dozens of experimental results accumulated over the previous night along with a cup of coffee. On the monitor in front of the entrance, the number of experiments processed that night and their success rate are updated in real-time numbers, and it is designed so that a warning alarm automatically sounds if the temperature or concentration falls outside a predetermined range.

When it was first revealed to the press in June 2026, the institute announced a plan to train 150 new personnel within this year, and several domestic pharmaceutical companies are already reviewing plans to bring this laboratory's operating method into their own companies. The sight that was

considered like a scene from a movie just a few years ago has now become a daily routine repeated every night in a laboratory in Seoul.

In 2024, the international academic journal *Nature* published a study in which an artificial intelligence based on a large language model controlled a spectrometer and a liquid handling robot by itself to synthesize twenty-nine types of organosilicon compounds without passing through human hands. Among these, eight types were compounds that had not been known to the world until then. The task that a chemist would have had to struggle with for days in front of a reagent cabinet, the artificial intelligence finished overnight by finding the reaction path by itself and coordinating the equipment schedule. In the past, discovering a new drug was a matter of tediously repeating over several years the four steps of finding a target that causes a disease, designing a compound that fits that target, synthesizing it, and confirming its efficacy with test tubes and cells.

This process, commonly called the DMTA cycle after the first letters of design, make, test, and analyze, was a slow circulation that took a person several weeks just to complete one round. The fact that the multinational pharmaceutical company Janssen changed the production process of the AIDS treatment Prezista to an automatic control system based on real-time sensors and machine learning to meet the strict standards of the United States Food and Drug Administration and the European Medicines Agency also lies on the same current. From a single robotic arm in Seoul to a page of an academic journal on the opposite side of the earth, the very definition of the act of an experiment is changing.

The force leading this change is agent AI, which thinks and acts by itself. It refers to an artificial intelligence that takes responsibility for making a plan and executing it when only a goal is given. In the past, medical artificial intelligence was closer to a passive tool that produced one output value for one input value. Agent AI is different. It continuously accepts data, remembers

previous judgments, and rewrites the next plan if it judges that the plan it set up by itself is wrong. Equipped with a multimodal capability to read vision, language, and chemical signals together, this artificial intelligence does not stay as software on the cloud but has risen to the level of directly commanding laboratory hardware like spectrometers and robotic arms.

An early architecture given the name ChemCrow demonstrates this principle well. Like a person calling various departments to divide and assign tasks, it is a method where a large language model directly issues commands to spectrometers and chromatography equipment through predetermined communication standards, so-called application programming interfaces. If a researcher only inputs the target substance and the safety conditions to be kept, the artificial intelligence searches through a vast chemical reaction database to map out a synthesis route, orders reagents, coordinates equipment schedules, and after executing the experiment, analyzes the results and fixes the next conditions by itself.

This flow, connecting everything into one without passing through human hands, from the task of reading literature to making hypotheses and synthesizing and analyzing with robots, is called a closed-loop system. Unlike existing automated equipment that only repeats a set sequence, this system is fundamentally different in that it restructures the procedure to be executed next by itself when the situation changes.

The first step of the closed loop is the task of making hypotheses. 'Coscientist', released by Google DeepMind, is a system where four agents divide roles and collaborate based on the Gemini large language model. When the generation agent skims papers and data to produce several hypotheses, the reflection agent digs out their flaws like a peer reviewer. The ranking agent makes the surviving hypotheses compete in a tournament format, and the evolution agent refines and combines the excellent hypotheses among them.

The four agents exchange each other's results and review the same hypothesis repeatedly, and this repetition is close to a structure mimicking the debates a person might have with colleagues for months ahead of an academic conference presentation. The research team of Professor José Penadés at Imperial College London in the UK had been studying for nearly 10 years how antibiotic resistance genes spread among bacteria, and they entrusted this yet-to-be-published research topic to Coscientist exactly as it was. Even without access to the research team's unpublished data, the artificial intelligence produced a hypothesis pointing to the same conclusion in just 48 hours.

The hypothesis verification process that human researchers used to establish and refute over several weeks to months, these four agents run automatically in a much shorter time. The problem is that the faster hypotheses come out, the more the hands to actually confirm those hypotheses emerge as a new bottleneck. The self-driving laboratory is the next step that appeared precisely to fill this bottleneck.

When a hypothesis arrives at the robot, the next step is physical execution. China's XtalPi operates roughly 300 robotic workstations around the world around the clock without stopping, combining roughly 500 proprietary AI models with quantum mechanics calculations to synthesize and evaluate compounds. The company is currently conducting joint research with Eli Lilly worth up to \$250 million, and domestically it is working with JW Pharma to refine lead compounds targeting STAT6, a target protein for inflammatory diseases. Molecular structures and protein sequences drawn by AI algorithms are immediately transmitted to robotic assembly lines and cell culture systems and transformed into physical substances.

Every product created in this process is assigned a unique number linking from the computer design to the final production batch and is tracked all the way through, enabling people to later retroactively trace which experimental

conditions led to which results. This also serves as at least a minimal safety net. The robotic arm at Dark Lab in Seoul handling reagents in 0.005-milligram increments and the fleet of 300 robots in a laboratory on the opposite side of the world both operate on the same principle. The more repetitive work that would be error-prone for humans, the more tireless the robot's hand remains, and the ability to repeat the same conditions identically a hundred or thousand times increases the value of this physical execution stage.

When the experiment ends, the cycle does not stop but moves to the next phase. Multiple sensors embedded in the bioreactor read in real time the gene expression of cells, electrical signals, and traces of metabolic products and send them to the AI control device. The reinforcement learning controller receives these signals and autonomously adjusts the concentration of the culture medium or the amount of signaling molecules, taking action before quality deviates from the standard. The data from result analysis returns to the initial hypothesis design stage and becomes material to refine the conditions for the next experiment more precisely. In this way, the four stages - hypothesis design, lab automation, result analysis, and data feedback - form a circle and repeat endlessly.

Yesterday's failed experiment becomes material for today's successful experiment, and today's success gives rise to tomorrow's question. Because the system itself becomes gradually smarter with each experiment cycle, this laboratory is not a static space that must be manipulated by humans but rather resembles a living organism that learns by itself and grows. Unlike the past when humans worked eight hours a day and handled a few experiments, now in the same time period hundreds of conditions can be tested simultaneously, and the results are immediately reflected in the next experiment, creating a difference in scale.

The potential of autonomous laboratories is being confirmed outside corporate press releases. Multiple AI agents forming a team designed

nanobodies that bind to the novel coronavirus - that is, small binding proteins modeled on antibodies found in camels and alpacas - and verified their effectiveness through actual experiments in what was called "Virtual Lab" research, which has been published in academic journals after peer review. A system called "Robin" created by America's Future House integrates the process of forming hypotheses, designing experiments, and interpreting results and newly identified the glaucoma drug ripasudil as a treatment candidate for dry age-related macular degeneration - that is, an eye disease that gradually causes vision loss with age.

This research was published in Nature, an international academic journal, in May 2026. In that both cases were validated by fellow researchers at universities and research institutes and remained as published papers, autonomous laboratories are not merely promotional language for attracting investment but support the fact that they are actually functioning scientific methodologies. Not only companies backed by large capital but also university research teams with relatively limited budgets are finding a path to join this trend through leasing AI in the cloud and shared laboratory equipment.

Bio venture capital Flagship Pioneering in March 2025 introduced to the world a company named Lila Sciences and announced that it received \$200 million in seed investment. In October of that year it additionally attracted \$115 million led by Nvidia's investment division, bringing cumulative investment to \$550 million and raising the company valuation to around \$1.3 billion, and in June 2026 news was received that follow-up investment discussions recognizing a company valuation of \$8.5 billion were underway. Liam Fedders, who served as vice president of research at OpenAI, and Ekin Douss Truuk, who worked as a researcher at Google DeepMind, together founded Periodic Labs and received \$300 million in seed investment in 2025, building autonomous laboratories targeting new materials including superconductors.

Chemify, founded by Professor Lee Cronin of the University of Glasgow in the United Kingdom, uses "chemputer" technology that converts chemical reactions into code and has robots reproduce them exactly, and received over \$50 million in Series B investment in 2025. All three companies are not targeting just a single new drug. What they are aiming for is the automation of research and development itself encompassing chemistry, materials, and energy, and the investor list also includes names such as Nvidia, Amazon founder Jeff Bezos, and former Google CEO Eric Schmidt. The investment funds accumulated by the three companies combined amount to \$900 million, and to a field that just a few years ago was considered roughly on the level of a research project in a university laboratory, the very fact that such funding is flowing in shows the weight of validation that this technology has passed.

The results produced by autonomous laboratories are already being confirmed by concrete numbers in several countries. Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO) has reduced the time it takes to develop a new catalyst from 5 years to 6 months. Canada's Accelerate Consortium discovered over twenty-one new candidate materials through autonomous laboratories. In Korea as well, the Korea Institute of Science and Technology (KIST) developed its own operating system called "Octopus" and is conducting experiments directing multiple robotic devices with a single AI, and the Korea Pharmaceutical Industry and Biotechnology Association and Yonsei University are separately building related infrastructure through the K-NIBRT project.

Both projects are still in their infancy, experienced by only a small number of researchers. However, they serve as a bridge to the next stage because they bring together the automated experiments that the government and universities had been trying separately under a single standard. Although their nationalities and research subjects are different, they share one common point: artificial intelligence and robots are completing in a few months, or sometimes a few weeks, the work that took people several years to do.

The South Korean Ministry of Science and ICT has announced a plan to build six autonomous laboratories with an investment of 49.5 billion won over three years from 2026 to 2028. Catholic University of Korea will take charge of one general-purpose laboratory to build the K-Cell platform. Five specialized laboratories will be assigned as follows: Daegu Gyeongbuk Institute of Science and Technology for liquid biopsy, KAIST for infectious diseases, Korea University for gene delivery vectors, POSTECH for enzyme engineering, and Ulsan National Institute of Science and Technology for organoid-based drug evaluation. Oh Dae-hyun, the Director-General for Future Strategic Technology Policy who leads this project, stated that autonomous laboratories combining artificial intelligence and robotics will be a turning point that fundamentally changes the paradigm of biological research and development.

Separately, the government has decided to invest an additional 34 billion won over two years to secure domestic technology for physical artificial intelligence, which corresponds to the brains of robots. The scene of overnight experiments that Darklab in Seoul has previewed is preparing to expand into a picture that unfolds simultaneously in six locations nationwide by 2028.

Some in the industry are proposing the idea of compressing the time it takes from reading a single paper to filing a patent to around 90 days. There are not yet any officially recognized cases of this actually being achieved, but numbers already exist that suggest this direction is not just an empty slogan. Insilico Medicine, headquartered in Hong Kong, stated that it takes an average of 12 to 18 months to find preclinical candidates, which are substances ready to enter animal testing, using its target discovery program PandaOmics and its new drug candidate design program Chemistry42.

This figure is less than half of the two and a half to four years that traditional drug discovery used to take. One program conducted by this company spent only 2.6 million dollars to find a new molecule for a new target without synthesizing even eighty candidates, which is one-tenth of the cost required by

conventional methods. The candidate substance for treating idiopathic pulmonary fibrosis introduced by this company was the first among drugs designed by artificial intelligence to have its effect of improving lung function confirmed in phase 2 clinical trials. In 2025, it advanced a new candidate, ISM9528, aimed at pain treatment, to the preclinical stage, throwing down a challenge in a field with an annual market size reaching 100 billion dollars.

A program called inClinico, which this company also operates, is a tool that uses data to estimate in advance the probability that a phase 2 clinical trial will proceed to phase 3 and succeed. Since the same company's artificial intelligence is involved from beginning to end, from the stage of finding candidates to the stage of designing clinical trials, it can be called a miniature version of a closed loop.

Quantum machine learning (QML), which combines the phenomena of entanglement and superposition from quantum mechanics with machine learning, calculates fine interactions between molecules with a precision that is difficult for conventional computers to handle. Qubit Pharmaceuticals in France used Orion, a neutral-atom quantum computer made by Pasqal, to calculate exactly where water molecules are positioned inside the binding pocket of a protein using quantum algorithms. In March 2026, the Cleveland Clinic in the United States and IBM announced that they successfully calculated the electronic structure of Trp-cage, a miniprotein made of 303 atoms, using a hybrid method that alternates between quantum and classical computers.

Quantum algorithms are providing more precise answers than conventional methods even in reading in advance subtle interactions created by the simultaneous entanglement of multiple atoms, such as whether a drug will cause unintended toxicity to collagen tissue inside the body. Market research firms forecast that the market size for this field will grow at an annual average of 17.9 percent, from 126 million dollars in 2025 to 638 million dollars in 2035.

This is coupled with continuous manufacturing technology, which keeps the process flowing without stopping from the input of raw materials to the production of finished products. Process analytical technology and digital twins detect and simulate in real time the minute temperature changes and impurities that occur during the drying, powder mixing, and tableting processes that press out pills. The case where Janssen applied such a system to its production line to meet regulatory standards is a representative example.

This means that the speed of discovery and the speed of production have now begun to move to the same rhythm. As candidates found in the laboratory in a few weeks can now meet the conditions for mass production in the factory in a few weeks as well, the gap of time that previously lay between discovery and production is narrowing itself.

The industry has given this integration the name Industry 5.0, going beyond automation. It means the next stage after Industry 4.0, where machines replaced human repetitive labor; it is a time when artificial intelligence and robots become companions that extend human cognitive and physical limits. The fact that blueprints inside a computer server connect seamlessly through laboratory robots to the production lines of a factory is also one of those results. As nights and weekends, which previously could not be reached by human hands, are incorporated into research and development, the very number of candidates that can be extracted with the same budget and the same personnel is multiplying several times over. However, even companions need rules to follow.

If you step inside this closed loop, you can see the path a new drug takes. In the target discovery stage, where disease-related proteins are found, artificial intelligence scans vast amounts of genomic and literature data to select strong targets. Once the three-dimensional structure of that protein is predicted, artificial intelligence draws a perfectly fitting molecule from scratch. At the

same time, it calculates in advance how that molecule will break down in the liver, how it will leave the body, and whether it is toxic. Sometimes, a single small compound is not enough. In those cases, the candidate is redesigned into a multi-target antibody or cell therapy that aims at several targets at once.

These drug candidates are refined once more so they can be delivered exactly to desired tissues, such as the liver or a tumor. They are loaded onto nanoparticles made of fat droplets or polymers. Organoids, which are mini-organs grown in a dish from a patient's cells, and digital twins, which are exact computer copies of a patient's whole body, serve as testing grounds. Here, the effects and side effects of the candidates are tested before animal experiments begin. The laboratory itself is no exception. Before a robot uses real chemicals, the experimental conditions are first run in a virtual laboratory on a computer. Only the conditions with a high chance of success are passed on to the physical robot. This method is becoming widely used.

The candidates that survive this process enter clinical trials designed in advance by digital twins. They then meet the tools of precision oncology, which read the genetic map of a tumor to find the right treatment for each individual patient. The records from all these steps are also gathered as drafts for the documents that will be submitted to regulatory authorities. From finding a target to finding the right prescription for a patient, tasks that used to operate separately in different departments now run at the same time. They exchange data within a single continuous loop.

This cycle does not always run smoothly. There is a real threat that an agent AI might design dangerous materials on its own due to biased data or errors. It could also apply the wrong conditions in a continuous manufacturing line, mass-producing defective medicines. Because of this, more laboratories are creating a separate "guard agent." This guard agent restricts the artificial intelligence's activities within pre-set safety limits and watches over it. If it detects an attempt to cross the set limits, the guard agent is designed to stop

the experiment immediately and hand the decision over to a human. This acts as a double safeguard where autonomy and safety keep each other in check.

Technical flaws often occur, such as data mismatches caused by synchronization errors between sensors and software. The black box problem also remains, where humans cannot fully explain why the artificial intelligence made a certain decision. To control these risks, the United States Food and Drug Administration and the European Medicines Agency published their first joint guidelines in January 2026 on using artificial intelligence in new drug development. The core principle is that a human must finally confirm the artificial intelligence's suggestions, especially for decisions that affect quality. Pharmaceutical companies must document the learning process and bias verification procedures of their artificial intelligence. This documentation must follow the revised computer system validation standard guidelines from the International Society for Pharmaceutical Engineering. It must also follow the action plan for applying artificial intelligence to software classified as a medical device, and the data integrity principle, which requires tracking exactly who created the data and when.

Artificial intelligence changes slightly as it learns on its own. To control this phenomenon, known as model drift, the United States Food and Drug Administration has introduced a predetermined change control plan system. Under this system, the range in which an artificial intelligence can safely update itself after it goes on the market is set and approved in advance.

Moving into the realm of law, the questions become even sharper. Can a patent be given to a molecule drawn from scratch by an artificial intelligence that read tens of millions of papers without any real human intervention? The answer varies from country to country, and a global standard has not yet been set. Suppose a dosage optimized by an autonomous laboratory causes unexpected side effects in a patient. There is still no legal system anywhere that clearly decides who is responsible: the technology company that wrote

the software, the pharmaceutical company that used it, or the regulatory authority that approved it. While the United States and Europe are creating guidelines at their own speeds and in their own ways, many countries, including South Korea, have not yet found their own answers to these questions.

If the decisions of patent offices and courts differ by country, a real problem could happen. A patent for the exact same molecule designed by the exact same artificial intelligence might be approved in one country but rejected in another. Technology crosses borders, creating twenty-nine new chemical compounds overnight. However, the legal language needed to take responsibility for those results is still moving at different speeds in each country.

My thought on what humans will do in a laboratory where robots replace hands is this: researchers will not disappear. However, their roles are fundamentally changing. The manual labor of moving chemicals and using scales will go to robots. Yet, deciding what questions to ask in the first place, and deciding which data to trust and which to doubt, must still be done by humans. A robot cannot replace the human eye that tells a coincidence from a true discovery among hundreds of results produced overnight by artificial intelligence. A robot cannot replace the judgment needed to draw the safety lines that a system must not cross. And a robot cannot replace the legal and ethical oversight needed to answer who is responsible when an accident happens.

Researchers appear to be moving from being technicians holding pipettes at a lab bench to being directors who design the entire laboratory and take responsibility for the mistakes that laboratory might make. Students studying chemistry or biology at universities are now also expected to learn the basics of statistics, coding, and robotics. It is likely for this same reason that the association operating dark labs has made training new talent its top priority.

Instead of an apprenticeship education where lab skills were learned by hand, the new basic skill for the next generation of researchers is the training to verify the decisions made by artificial intelligence and to push past its limits.

Even at night when the lights in the dark lab are turned off, someone had to give the robot its orders on what to experiment that day. And the next morning, there must still be a designated person who will sign off on the results.

In 2028, when six of Korea's autonomous laboratories funded with 49.5 billion won open as planned, a scene will unfold where experimental laboratories across the country, together with Seoul's DarkLab, simultaneously turn off their lights every night and operate robots. On the other side of the Earth, CrystalPy, which operates 300 robots, and Raila Sciences, which has been recognized with a company valuation of 8.5 billion dollars, will be continuing experiments at the same time in the same way. Periodic Labs' robots aiming at superconductors and ChemiPy's chemputer translating chemical reactions into code will also work through the night using the same principles while targeting different substances.

This cycle of reading literature, forming hypotheses, synthesizing materials, and feeding back results does not stop even while people sleep. The AI that finds targets, the AI that draws molecules, and the control device that moves the robot arm now work together like a single cycle, scattered as they are across different companies and different countries. However, when one early morning an unexpected number emerges from one configuration that an autonomous laboratory has optimized overnight, who bears responsibility for it is not yet clearly written in any country's legal code.

The robot does not provide an answer to that question. Instead, that question is quietly placed on the desk of the researcher who arrives at work the next morning, along with dozens of pieces of data accumulated overnight. The task of preparing an answer to that returns to humans every morning when the

laboratory lights come back on.

The autonomous laboratory A-Lab, created by the United States' Berkeley National Laboratory, published a paper in *Nature* in 2023 stating that robots discovered dozens of new inorganic compounds over seventeen days without human intervention. A few months later, external researchers including Robert Palgrave, a chemist at the University of London, took another look at those results. Among the substances said to have been 'newly discovered,' many were compounds that were already registered in the crystal structure database - that is, in data collecting the atomic arrangements of known substances. *Nature* published a correction in January 2026.

The Berkeley research team leading this laboratory refuted this, claiming that A-Lab had indeed actually synthesized those compounds, and the debate has not yet ended. This incident appears to reveal the fact that the more the entire process from forming hypotheses to executing experiments to interpreting results is handled by a single AI in a closed loop, the fewer outside eyes there are to reexamine that judgment. Whether the next generation of researchers who learned chemistry only through numbers on screens without the experience of directly handling reagents will be able to catch such errors is something no one can yet guarantee.

There was already an instance where a safeguard against AI designing dangerous substances was put to the test. In 2022, researchers at the American pharmaceutical company Collaborations Pharmaceuticals flipped the objectives of their drug design AI MegaSyn, which had been trained to avoid toxicity, and ordered it to find molecules that would maximize toxicity. Running on just one ordinary laptop for six hours, the AI generated a list of as many as 40,000 toxic substances, including the nerve agent VX prohibited by the Chemical Weapons Convention.

Rather than publicly releasing these results in detail, the researchers reported the danger to *Nature* Machine Intelligence while keeping the specific chemical

structures confidential. The uncomfortable lesson this experiment leaves behind is that there is no difference between an AI trained to make new drugs and an AI trained to make poison other than a change of objectives made by a person.

All that was needed to reverse MegaSyn's objectives was changing a few lines of code. In an era when robot arms weigh reagents on scales all night and AI decides the next experiment on its own, there must be someone inside the laboratory to notice that gesture and stop it. If someone places before you tomorrow morning a drug that an AI designed overnight, there is a question you must ask before asking whether that drug was made by a person or a machine. It is the question of who verified that drug and by what method.

The Paradigm Shift in AI Drug Discovery and the Future Convergence Ecosystem

Author | Kim Kyung-jin

Publisher | Kim Kyung-jin

Imprint | Kim Kyung-jin Attorney Press

Price : USD 15

© Kim Kyung-jin 2026

All rights reserved. No part of this book may be reproduced without permission.

Note) AI tools assisted with parts of the research and editing of this book.

kimkj008@gmail.com

Support This AI Library Book

If this book was useful, you can support future translations, public editions, and open AI knowledge projects through PayPal.



PayPal

<https://paypal.me/kimkjkj>

Scan the QR code or open the link above.