



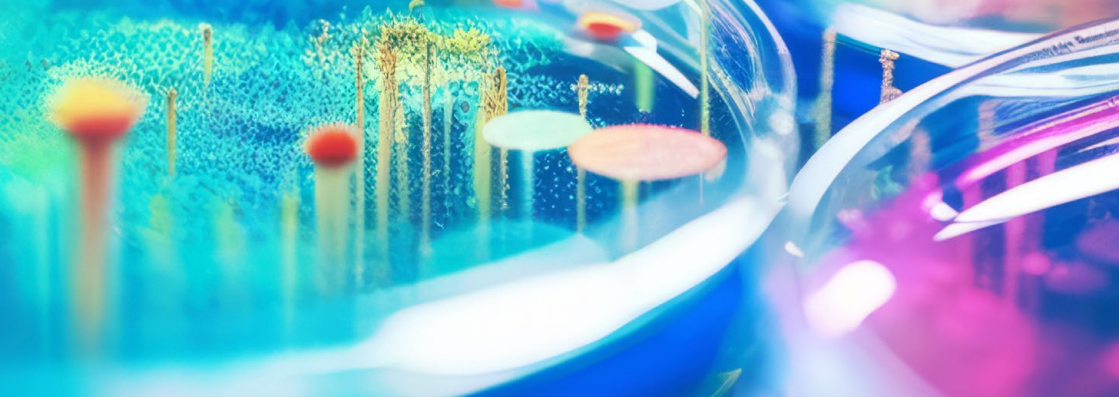
The AI research assistant

**A Thoughtworks methodology
for accelerating healthcare
and life sciences R&D**

/thoughtworks

Strategy. Design. Engineering.

Capabilities and data grow stronger every day. But R&D is getting harder to optimize	3
Knowledge management: A major challenge in R&D	6
The AI R&D opportunity	8
Introducing the Thoughtworks R&D digital assistant	11
Three big ways the R&D digital assistant can transform knowledge management	13
How we built the R&D digital assistant	17
Accelerating drug discovery during preclinical research at Bayer	21
Enabling R&D teams to talk to their data products at Pfizer	23
Three steps to AI-enabled R&D transformation	25
Transform knowledge management and unleash R&D productivity and efficiency	27



Capabilities and data grow stronger every day. But R&D is getting harder to optimize.

Research and development (R&D) is the lifeblood of the healthcare and life sciences (HCLS) industry. It's a continuous process that provides competitive advantage, helps solve major societal challenges, and pushes the boundaries of what's possible for modern medicine.

The past decade has brought immense change for R&D teams. Major technology, data and science advancements have empowered teams to achieve more than ever before. Teams have invested heavily in those capabilities, with total pharmaceutical R&D spend set to exceed \$200 billion by 2025¹.

But despite those investments, many R&D organizations saw their productivity fall steadily between 2010 and 2019 — and levels have fluctuated strongly ever since².

1 <https://clarivate.com/innovation-exchange/solution/cmr-international-pharmaceutical-rd-factbook/>

2 <https://www.deloitte.com/uk/en/Industries/life-sciences-health-care/research/measuring-return-from-pharmaceutical-innovation.html>



It now takes between 10 and 12 years and between \$161 million and \$4.5 billion for an FDA-approved drug to reach the market.³

While data-driven workflows, tools and capabilities have transformed how R&D teams work and enabled them to evolve how drugs and therapies are discovered and brought to market, they've also created major challenges.

Managing huge volumes of complex data

Life sciences R&D generates massive amounts of data, including genomic, proteomic, metabolomic and clinical data from multiple sources, which are complex and difficult to manage.

Upholding FAIR data standards

The complex and evolving nature of data due to growth and consumption patterns restricts standardized data structures and models and makes it difficult to implement FAIR data standards⁴.

3 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8516790/>

4 <https://www.go-fair.org/fair-principles/>

Ensuring strong data privacy and security

Life sciences R&D involves sensitive personal health data, making data privacy and security paramount. This requires implementing robust data governance and security measures to protect against data breaches.

Enabling secure data sharing and collaboration

Collaboration between research groups and organizations is essential for advancing life sciences R&D, but managing data governance across different organizations can be challenging.

Analyzing and interpreting data at scale

Gleaning insights from large, complex datasets requires specialized skills and expertise. R&D teams must have access to appropriate data analysis tools and resources to extract meaningful insights from their data.

That's a heavy laundry list of challenges. But many of them come together to create one overarching issue for R&D teams: **Efficiently creating, retaining, sharing and leveraging knowledge across the organization.**



Knowledge management: A major challenge in R&D

HCLS organizations hold vast quantities of data and knowledge that can be applied to accelerate R&D processes and help teams use findings from previous studies in ongoing projects. However, complex data and system environments make much of that knowledge difficult — or virtually impossible — to leverage at the right time.

How often do you hear questions like these from your R&D teams?

**"Has this observation been documented
in previous Phase 1 studies?"**

**"Are there existing clinical studies that explain
my current observation?"**

"Can I generate an FDA compliant study report?"

**"What were the critical decisions
made for this molecule?"**

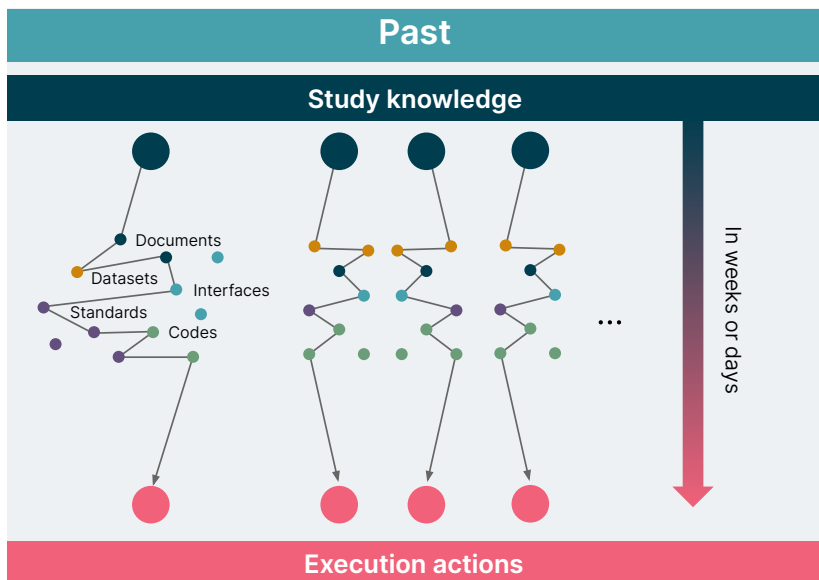
“Can I use R code for immediate data visualization and analysis?”

“Can I quickly search through extensive institutional knowledge to accelerate decision-making?”

Valuable knowledge exists within these organizations, but teams simply can't find and use it when they need it, or efficiently convert their knowledge into new, value-creating actions.

The AI R&D opportunity

In the past, documents, datasets, interfaces and other valuable knowledge assets have been stored in fragmented systems — some highly structured in data warehouses, and some in unstructured formats like documents, PDFs and presentations. So, when R&D teams need to use them, it typically takes days or even weeks to surface that knowledge and translate it into new execution actions.

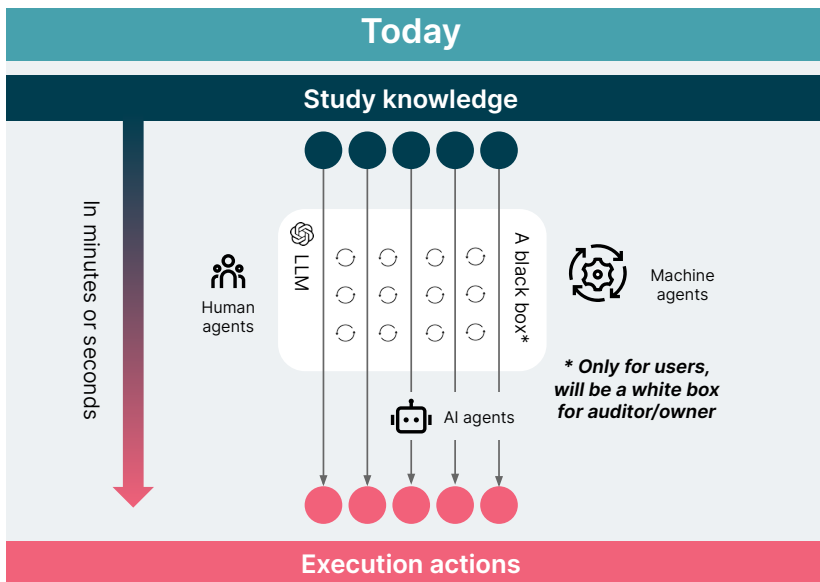


The previous path from study knowledge to execution actions

It's a long and frustrating process. First, teams must manually uncover what relevant knowledge the organization holds. A lack of visibility means that, in many cases, valuable knowledge goes unutilized because teams don't know it's there.

When they do find the data, they need to make sense of it and correctly interpret it both in the context of its original creation, and in the context of the current study they want to apply it to. Again, different knowledge formats and structures make this a tricky process that can slow R&D processes significantly.

But, by applying AI in the right ways, organizations can fundamentally transform how they convert past study knowledge into new execution actions.



How study knowledge converts to new execution actions today

In this diagram, we can see how a combination of human and machine agents use a black box AI solution to streamline knowledge search and interpretation processes. This cuts a process that previously took days or weeks into one that can be completed in minutes or even seconds.

Let's take a closer look at exactly how that scenario translates into a deployed solution.



Introducing the Thoughtworks R&D digital assistant

Over the last few years, Thoughtworks has helped numerous major HCLS organizations overcome persistent challenges around knowledge retrieval, interpretation and creation and optimize R&D cycles using AI-enabled digital solutions.

Throughout our work, we noticed several patterns — most notably, that common antipatterns in R&D cycles could be solved through the application of AI-enabled components. Now, we've compiled those components into a solution that any HCLS organization can use to transform how their R&D teams operate.

The Thoughtworks R&D digital assistant is a multi-agent platform within an organization's ecosystem. It's structured around use case-driven business agents that can be consumed individually or together across the R&D value chain.

The solution enhances productivity and innovation for R&D users by streamlining the journey from data to information, knowledge, actions and insights — minimizing the time and effort spent on knowledge discovery, analysis and creation activities.

Built alongside some of the world's largest and most successful HCLS R&D teams, the Thoughtworks R&D digital assistant is engineered to:

Improve and enhance data experiences for R&D teams

and make it easier than ever for teams to turn knowledge into valuable actions that improve the quality of everything they deliver.

Accelerate results and returns through robust knowledge management

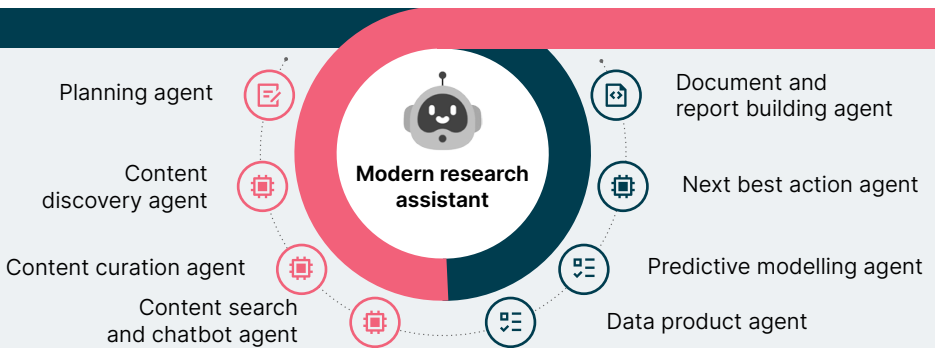
and ensure every new research project is strengthened by knowledge gathered in the past, creating a virtuous cycle of continuous improvement.

Support and augment activities across the R&D value chain

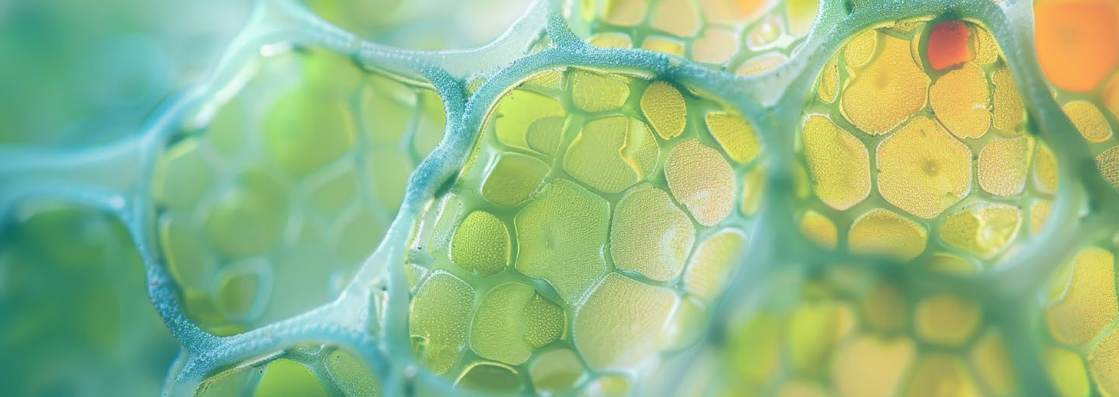
and free teams from the time-consuming manual tasks that prevent them from focusing on value-adding activities.

Empower teams at every stage of the R&D journey

The Thoughtworks R&D digital assistant provides HCLS organizations with AI-powered research assistants that can support them and streamline workflows at every stage of the R&D process. From surfacing relevant knowledge during planning, to authoring new content at the end of a project, it solves both the symptoms of poor knowledge management and its underlying causes.



The AI-powered research assistants that streamline workflows at every stage of the R&D process.



Three big ways the R&D digital assistant can transform knowledge management

The R&D digital assistant helps teams leverage AI to transform three core pillars of the R&D process:

Knowledge retrieval (ask)

The R&D digital assistant can help teams improve the accessibility, discoverability and searchability of knowledge by optimizing and accelerating search processes, and rapidly scouring diverse structured and unstructured data sources. AI search capabilities can be personalized at the team or individual level, prioritizing knowledge that's highly relevant to the searcher's needs.

The assistant effectively enables R&D teams to 'talk' to their data. They can directly ask for what they need in natural language. Instead of combing through numerous systems, their knowledge search begins and ends with a question, just like those you hear in R&D departments every day, such as:

- Were any of the following clinical findings observed in study T123456-2? Piloerection, ataxia, eyes partially closed?
- What were the clinical observations noted in trial T123456-2, particularly regarding the presence of hair bristling, impaired balance, partially shut eyes or soft bowel movements?
- What animal was used in study T-123456-2?

The key retrieval challenges solved by AI

Challenge #1

Data fragmentation and inconsistency across internal and external systems.

Challenge #2

Maintaining compliance with FAIR principles when using data from external sources.

Challenge #3

Poor searchability of unstructured document and knowledge formats.

Knowledge interpretation (analyze)

After teams identify knowledge relevant to their current studies, they need to interpret contextualized insights from it — a complex, time-consuming and resource-intensive process.

The R&D digital assistant can help transform how R&D teams interpret knowledge. Its immense processing power and generative capabilities enable teams to:

- Perform complex data analysis tasks at speed and make knowledge actionable faster.
- Spot patterns in millions of data points and gain new insights that human teams may struggle to deliver manually.
- Verify and validate knowledge at scale to ensure that the data they're interpreting is trustworthy and valuable to the business.

The ability to directly talk to data is also extremely valuable at this stage of the R&D process. With the help of a digital assistant, teams can simply ask AI questions like:

- Can you overlay the weight versus dosage data of study X with that of study Y?
- For the same studies, can you pinpoint all outliers?
- Can you find studies where similar outliers have been seen before?
- Can you detail out the conditions under which these studies were performed?
- What do all these studies have in common with each other that is not common with study Z?

When a team member uses the R&D digital assistant to ask natural questions like those about the knowledge they need, they can immediately access clear next best actions for their studies.

The key interpretation challenges solved by AI

Challenge #1

Dependency on computational scientists to access study reports and relevant analysis.

Challenge #2

Disparate analysis makes it difficult to integrate knowledge from various models and research approaches.

Challenge #3

Steep learning curve of advanced analysis and visualization tools.

Knowledge creation (Create)

Many of the knowledge management challenges faced by HCLS R&D teams are rooted in their knowledge creation practices. The manual nature of tasks such as report writing, study planning and content curation makes them extremely time consuming, leading them to frequently become an afterthought while research is ongoing.

The R&D digital assistant can help teams automatically capture knowledge and author content around it, removing manual reporting and recording workloads. Predictive modeling even enables teams to test hypotheses during the research process and record valuable findings before the end of a research project.

For existing knowledge, AI can help turn vast quantities of structured and unstructured insights into reports and documentation tailored to the needs of internal and external users.

The key creation challenges solved by AI

Challenge #1

Embedding standards for new knowledge creation across the R&D lifecycle.

Challenge #2

Recording knowledge in contextually relevant ways based on the original research questions asked.

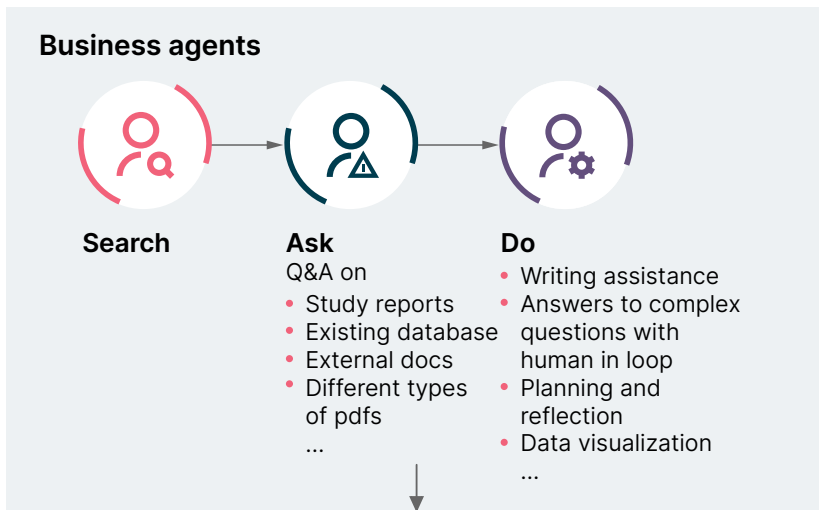


How we built the R&D digital assistant

The R&D digital assistant combines Thoughtworks' leading data and AI capabilities and agile engineering practices with robust, scalable cloud infrastructure to deliver a truly next-gen solution for HCLS organizations.

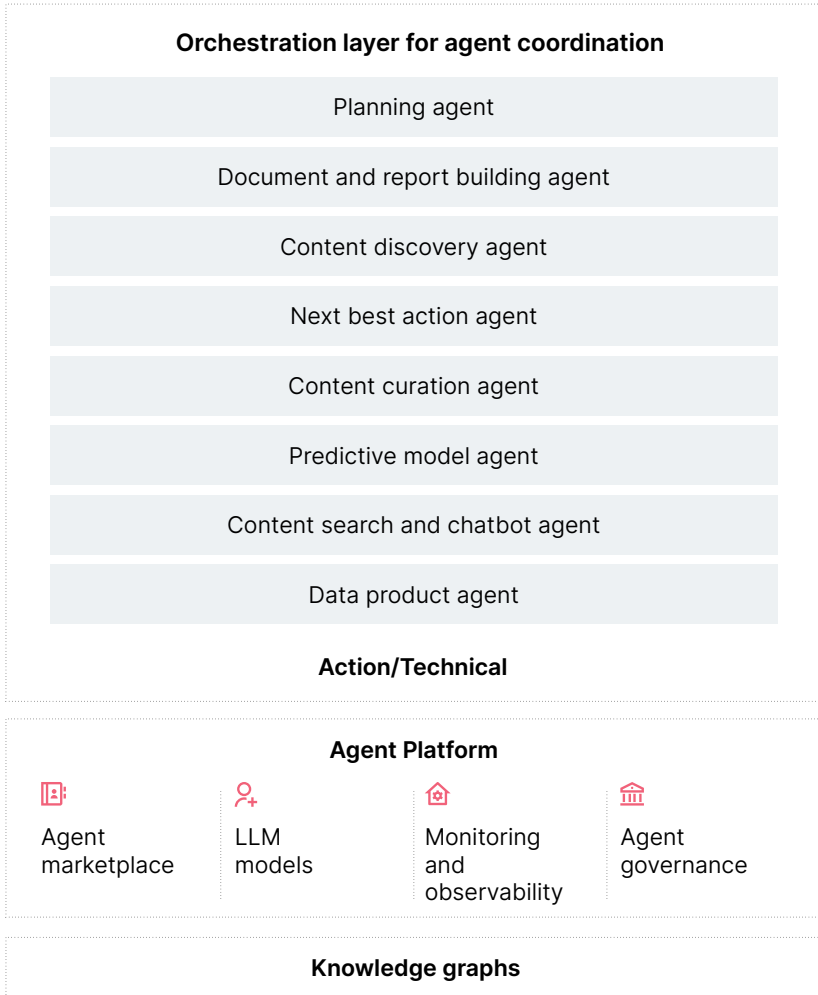
The solution provides a flexible, scalable platform that can rapidly adapt to changing market needs and regulatory demands. The platform enables real-time processing, automation and integration with third-party services — enhancing operational efficiency.

Here's a look at how it all comes together.



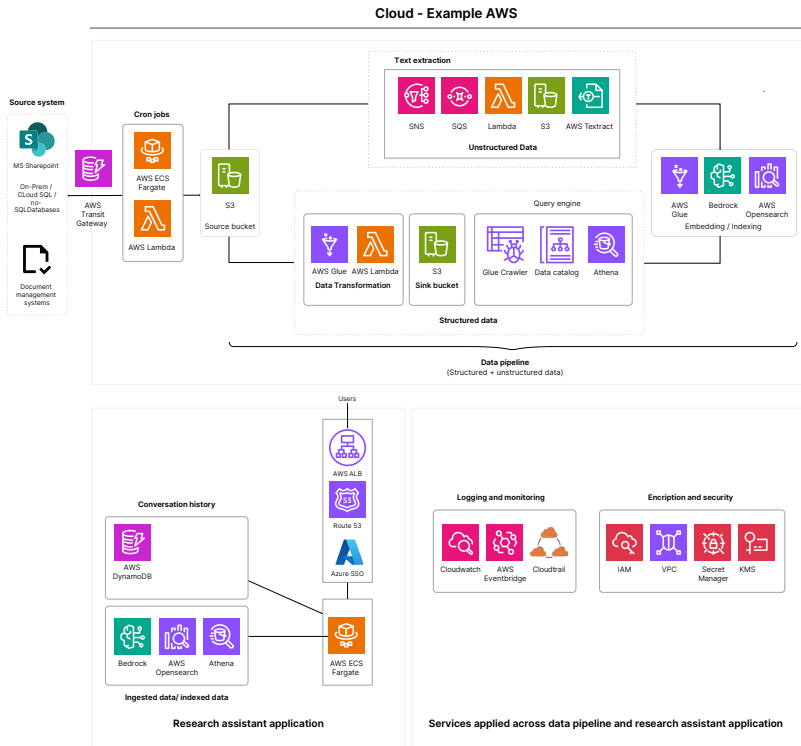
Foundational agents

(Used across multiple Business Agents)



An overview of the key elements and stages of the R&D digital assistant

The AI research assistant



A chatbot system built on a Retrieval-Augmented Generation (RAG) framework

The business agent consists of two main elements:
a data pipeline foundation and AI chatbot applications.

The data pipeline unites diverse data sources to bring structured and unstructured knowledge together and make it highly accessible and discoverable for R&D teams. Cloud capabilities including serverless orchestration, blob storage and NLP tools are used to catalog knowledge from PDF documents and extracted from other documentation formats so that all knowledge can be indexed together.

Similar capabilities do the same for structured data to create a common knowledge index that can be accessed and searched by AI chatbots. Serverless capabilities provide the crucial link between AI chatbot backend services and our indexed data pipeline, enabling the two to interface with ease.

Logging and monitoring, and encryption and security layers are applied across the chatbot application, to ensure all actions are carefully monitored to enforce responsible use of AI, and all data used by AI is kept secure and not exposed to any additional risk.



Accelerating drug discovery during preclinical research at Bayer

As one of the world's largest pharmaceutical and life sciences companies, Bayer's researchers and data scientists have immense volumes of past study data to leverage during their drug research processes. But previously, teams struggled to efficiently find critical insights from the vast number of past studies hidden deep inside unstructured PDF documents.

Thoughtworks and Bayer developed a Retrieval Augmented Generation-based AI chatbot to address these challenges. The chatbot enables users to query structured and unstructured data through an intuitive interface, providing rapid access to insights from study reports, summaries, conclusions and metadata.

Using our thin-slice approach, we first focused on delivering core functionalities to teams and then iterated on the solution based on user feedback. By following this approach, we were able to deliver critical capabilities to teams at speed. We could then shift our focus to the development of more complex new solutions, including a writing assistant which will help researchers generate first drafts of compliance-critical documents.

The chatbot has significantly improved research efficiency, enabling users to find highly specific insights from past studies,

plan new studies, onboard new team members and identify potential drug candidates more effectively. It's also reduced the time spent on document analysis and facilitated complex comparisons — accelerating the drug discovery process.

“I’ve been using the PRINCE Chatbot for a few weeks and it’s fantastic! It quickly finds studies, summarizes them and extracts the main conclusions. It also allows complex comparisons. For example, I used it to find a study on DNA strand breaks in female rats with higher toxicokinetic exposure, which greatly aided in planning my current genotoxicity study.”

Bayer R&D team member

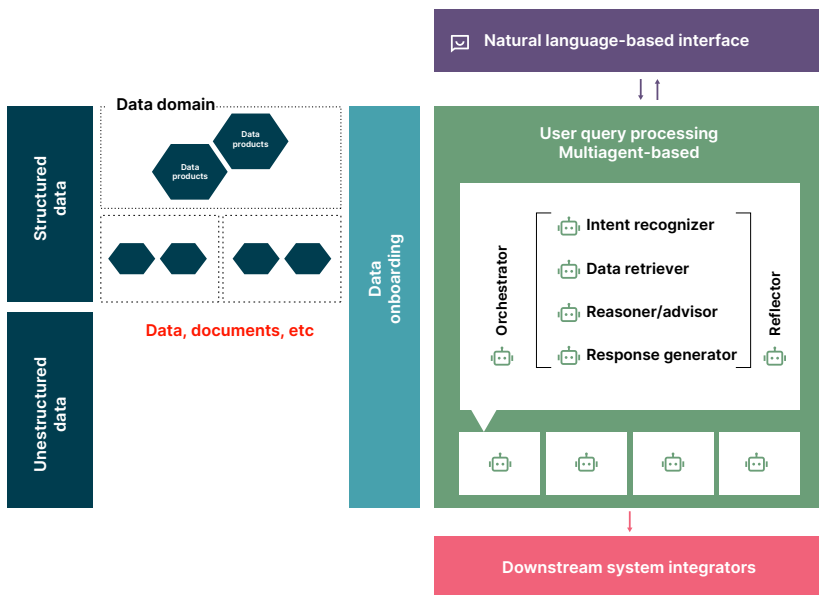
Empowered with an intuitive AI chatbot that connects research teams to relevant knowledge at speed, Bayer has seen:

- Significant time savings across the drug discovery process, resulting in new drugs being brought to market faster.
- Less time and effort spent on manual document analysis and complex comparisons.
- Greater utilization of knowledge across R&D teams, with fewer cases of research overlap and duplicated effort.



Enabling R&D teams to talk to their data products at Pfizer

At Pfizer, another major pharmaceutical company, we built an AI-driven research assistant capable of enhancing productivity and innovation by facilitating easier access to insights held within a defined set of data products.



The AI-driven research assistant we built at Pfizer

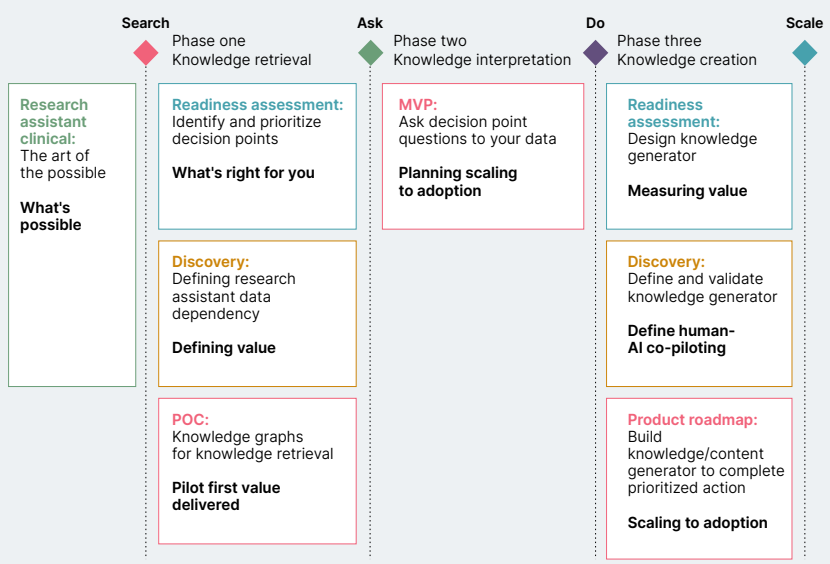
Using the natural language interface, biopharma research teams can now access relevant knowledge immediately, without the need for specialist data engineering skills. When a user asks a question using natural language, the solution streamlines the discovery, summarization and visualization of oncology data, immediately presenting them with the knowledge needed to expedite research and deliver better results.



Three steps to AI-enabled R&D transformation

When implementing your own R&D digital assistant, Thoughtworks recommends a multi-phase identification, design and building approach:

High level engagement model - research assistant



The multi-phase identification, design and building approach we recommend

After an initial exploration session to identify what might be possible for your organization, we recommend you first focus on knowledge retrieval. During this phase, your teams should:

Conduct a readiness assessment where they identify high-value decision points for R&D teams that could be supported with AI.

Undergo a data discovery exercise to define data dependencies and data product requirements for any research assistant use cases you may implement.

Build out knowledge graphs for knowledge retrieval and use them to create pilots and demonstrate value creation as early as possible.

With that foundation established, you can begin building MVPs that enable you to start asking questions of your data. That will then enable you to effectively plan how your digital assistant should be scaled out for adoption across R&D teams.

Finally, once you reach the third stage in the chart above, you can begin defining your human-AI co-pilot requirements, and build knowledge and content generators to complete your prioritized actions.

Transform knowledge management and unleash R&D productivity and efficiency

The R&D digital assistant by Thoughtworks has been developed through numerous engagements with leading HCLS R&D teams and provides a proven, easy-to-implement solution to the biggest knowledge retrieval, interpretation and creation challenges.

Through our customized implementation process, we identify the right combination of proven platform elements, bespoke capabilities and AI use cases with the potential to transform how R&D teams work.

Our solution helps high-performing HCLS R&D teams:

- Spend less time searching for knowledge and study data.
- Make informed decisions at every stage of the R&D lifecycle.
- Understand and interpret data in the context of its creation.
- Foster more effective research and improve scientific productivity.
- Utilize all the valuable structured and unstructured data they have access to.
- Generate stronger hypotheses at the beginning of new projects.
- Solve knowledge retention issues and separate knowledge from individuals.

- Create well-structured knowledge content automatically.
- Transform how teams engage with knowledge, and empower everyone to ask questions of their data and receive immediate, actionable answers.

That all adds up to huge productivity and efficiency gains for HCLS organizations — enabling them to innovate faster and remain competitive in an increasingly crowded and complex environment.

To find out more about the R&D digital assistant, or to explore how it could help you increase R&D productivity and efficiency, visit [thoughtworks.com](https://www.thoughtworks.com) or talk to us today.



Ammara Gafoor

**Healthcare and Life
Sciences Domain Lead,
Thoughtworks Europe**



Thoughtworks is a global technology consultancy that integrates strategy, design and engineering to drive digital innovation. We are over 10,500 Thoughtworkers strong across 48 offices in 19 countries. For 30+ years, we've delivered extraordinary impact together with our clients by helping them solve complex business problems with technology as the differentiator.

[thoughtworks.com](https://www.thoughtworks.com)



Strategy. Design. Engineering.

© Thoughtworks, Inc. All Rights Reserved.