SCITARIS
BIOPHARMA STRATEGY

Scitaris original research

A **biopharma-specific AI agent** for high-stakes strategic decision-making

Evaluating the performance of *Scitaris-AI* for in-licensing opportunity assessment and beyond

Large language models (LLMs) promise to transform strategic decision-making in the biopharma space and beyond – but can they be trusted? In our previous benchmark study[1], we demonstrated that commercially available LLMs – including Copilot, Claude, and Gemini – consistently misjudge pharmaceutical in-licensing opportunities, with error rates between 64% and 77%. While these foundation models showed promise in obvious cases, they systematically failed to incorporate nuanced details necessary for strategic decisions.

Here, we present *Scitaris-AI*, our approach to addressing these fundamental limitations through systematic integration of domain expertise, curated data sources, and structured analytical frameworks. *Scitaris-AI* achieves up to 95% accuracy and 94% reproducibility when evaluating defined strategic dimensions, working as a reliable agentic framework for screening of large opportunity sets.

Rooted in Scitaris' deep expertise and proprietary data, *Scitaris-AI* is being deployed as a configurable companion system to support deeper and broader strategic analyses across diverse biopharmaceutical business development applications.

1. THE PROBLEM:
Root causes of LLM unreliability

Our benchmark study revealed three fundamental failure modes rendering off-the-shelf LLMs unsuitable for high-stakes pharmaceutical decision-making[1]. First, **data quality issues** emerged through critical information gaps (models missed relevant specialized database content), hallucinated data (fabrication of non-existent clinical results), and indiscriminate source treatment (equal weighting regardless of scientific rigor). Second, **inferential and reasoning constraints** manifested as an inability to extrapolate from partial data, lack of structured analytical frameworks for multi-dimensional evaluation, and absence of industry-specific contextual knowledge. Third, **systematic biases in scoring** showed avoidance of extreme ratings (clustering toward neutral scores), optimism bias (39% false positives for weak opportunities), and uncorrelated assessments, which indicate randomness rather than systematic analysis.

These limitations stem from LLMs lacking validated data access, structured evaluation frameworks, and domain-specific decision criteria that expert strategy consultants employ. At this stage, automation – replacing human judgment with algorithmic output – remains unrealistic for complex pharmaceutical strategy. The solution requires systematic augmentation[2] – using AI as a collaborative partner – rather than sole reliance on model-inherent capabilities.

Background & unmet need summary

Off-the-shelf LLMs lack three critical capabilities for strategic pharmaceutical decision-making:

1. Access to high-quality, validated domain-specific data
2. Structured frameworks for systematic multi-dimensional analysis
3. The ability to apply industry-specific context and thresholds that distinguish promising from marginal opportunities

2. THE SOLUTION:
Augmented intelligence architecture

To address these limitations, we developed an agentic intelligence architecture that systematically integrates domain expertise with base AI capabilities (**Figure 1**). Rather than inventing new analytical approaches or blindly using AI's inherent capabilities, we translated Scitaris' established and validated consulting methodology, honed over years of strategic engagements, into a machine-executable framework. This provides LLMs with the data, protocols, and constraints that have proven reliable in expert analysis. In essence, we are treating LLMs like new hires: rather than expecting them to perform independently, we train them on how to conduct analysis – which data sources to consult, which frameworks to apply, and which criteria distinguish strong from weak opportunities.

2.1 Curated knowledge integration

The foundation of our approach is systematic integration of validated, proprietary domain-specific knowledge – a core differentiator of Scitaris that predates our AI architecture development. **Our pharmaceutical intelligence databases provide exclusive access to information unavailable through public sources.** This proprietary knowledge base is curated and continuously updated, undergoing rigorous validation to reflect the latest market dynamics, regulatory developments, and competitive landscapes.

In *Scitaris-AI*, the agentic framework queries not only defined public sources but also preselected internal databases, ensuring access to accurate, relevant, and structured information. This mitigates the data gaps and hallucinations observed in baseline models while providing informational advantages beyond what any public-facing LLM can access (**Figure 1A**).

2.2 Structured analytical protocols

Here, we show the implementation of reproducible analytical workflows for **three proof-of-concept dimensions critical for in-licensing decision-making: asset status, transactability, and competitive differentiation.** These dimensions were chosen to demonstrate workflow applicability across varying levels of complexity and types of data – from asset status as a fundamental and straightforward indicator, through transactability as a financial metric, to competitive differentiation as a multi-layered scientific assessment.

Each of the analyses follows strict protocols aligned with Scitaris' standard operating procedures (SOPs):

- **Development status** determines whether an asset is under active development by systematically evaluating multiple activity indicators (e.g., company operational signals, development communications, clinical trial registry status, ownership changes).
- **Commercial transactability** assesses out-licensing likelihood by applying tiered disqualifying factors based on recent business activity (e.g., new partnerships, financing rounds, ownership structure changes).
- **Competitive differentiation** evaluates whether an asset offers meaningful advantages over existing or advancing competitors (e.g., mechanistic differentiation, strategic positioning).

Each dimension employs precise decision criteria mirroring expert assessment, with explicit rules for weighting evidence. **This eliminates arbitrary scoring patterns and mirrors the strategic assessment by PhD-level experts** that has defined Scitaris' world-leading consulting practice.

2.3 Tailored configuration

By inputting authoritative databases, specifying exact frameworks, and tuning hyper-parameters of LLM function (e.g., minimizing randomness by lowering temperature[3]), we aim to ensure accurate and reproducible analysis (**Figure 1B**).

Critically, **our architecture can be tailored to specific client needs and strategic priorities.** Differentiation criteria, competitive benchmarks, and transactability thresholds can be customized based on a company's therapeutic focus, geographic priorities, deal structure preferences, and portfolio gaps. This flexibility ensures alignment with each client's unique strategic objectives while maintaining analytical rigor.

Architectural principle

Our augmented intelligence architecture does not attempt to make foundation models smarter about biopharma strategy. Instead, it systematically provides them with the specific knowledge, structured reasoning frameworks, and validated data that human experts use – enabling reliable analysis and maximizing LLMs' inherent capabilities while being model-agnostic. In this way, the architecture can support PhD-trained consultants to execute broader, deeper biopharma-specific strategic analyses without additional time requirements.

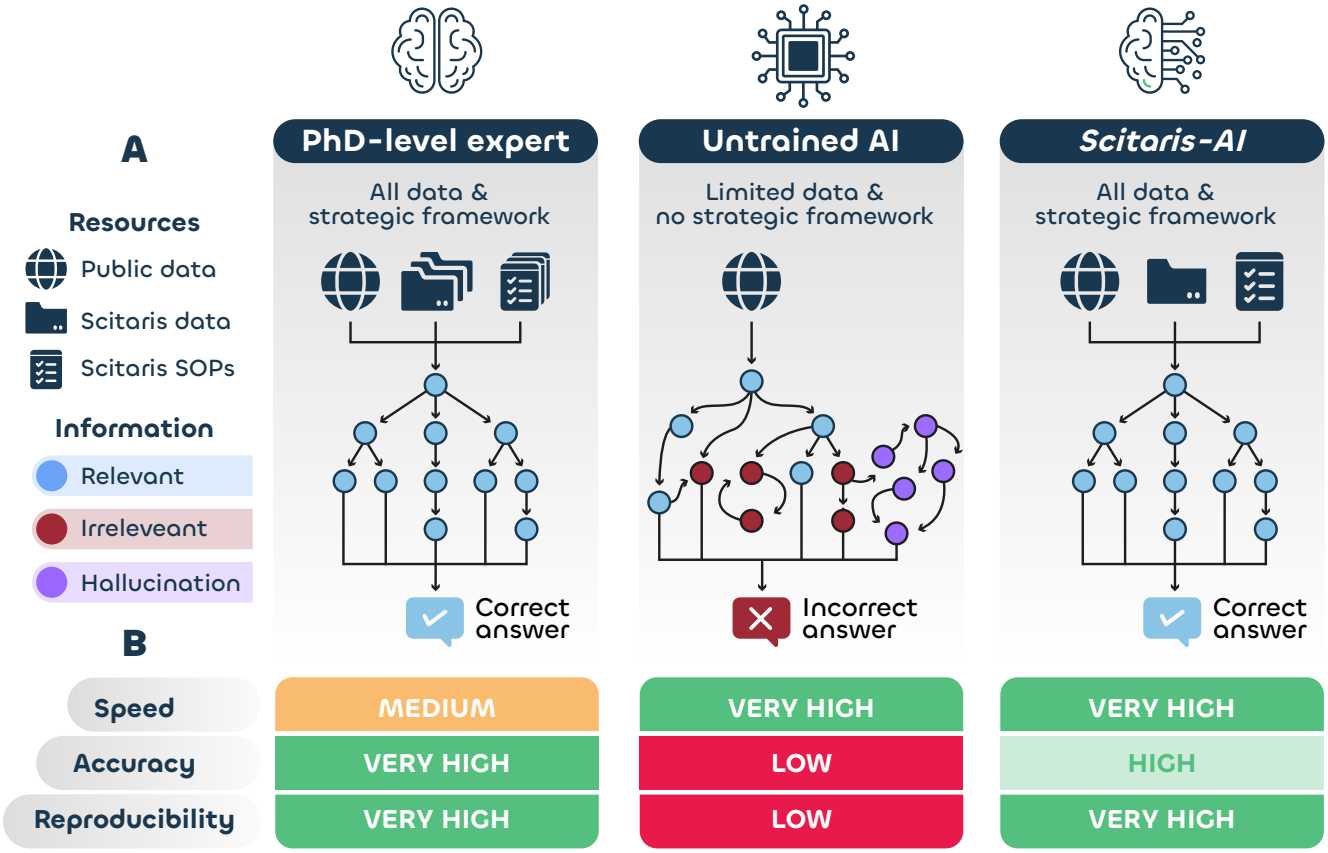

Figure 1. Comparison of strategic assessment approaches by PhD-level experts, untrained AI, and *Scitaris-AI*.
A. Resources and analytical pathways. Expert PhD-level consultants access comprehensive data sources and follow consistent pathways to correct answers. Untrained AI relies solely on public data, exhibit variable reasoning with hallucinations, and produce unreliable outputs. *Scitaris-AI* combines curated databases with explicit frameworks to achieve consistent, accurate results. Icons indicate data sources: public data (globe), proprietary databases (folder), and standardized operating procedures (checklist). Decision nodes show information handling: relevant (blue), irrelevant (red), and hallucinated (purple) data.
B. Operational characteristics across approaches. PhD-level experts deliver very high accuracy and reproducibility at intermediate speed. General untrained AI offers very high speed but low accuracy and reproducibility. *Scitaris-AI* achieves high accuracy and very high reproducibility at very high speed, combining the reliability of expert analysis with substantial efficiency gains that can support broader and deeper analysis within a given timeframe.

3. BENCHMARKING:
Accuracy and reproducibility

To validate our approach, we systematically evaluated *Scitaris-AI*'s accuracy and reproducibility in a proof-of-concept in-licensing database consisting of a set of 199 pharmaceutical assets previously assessed manually by our team. To assess accuracy, we applied *Scitaris-AI* across the three selected dimensions for each asset and compared the outputs against expert ground truth assessments (**Figure 2A**).

***Scitaris-AI* demonstrated 89-95% accuracy**, representing dramatic improvement over the 23-36% accuracy rates observed with off-the-shelf foundation models[1]. Critically, **the percentage of false negative results was low**, ranging between 2.9% and 4.6%. To assess reproducibility, we ran the same assessment across three independent runs and compared the results; when all three runs reached the same outcome, the assessment was considered reproducible. **The workflows demonstrated high reproducibility:** >93% of assets received identical classifications across the three runs for status and transactability, and 83.4% for differentiation (**Figure 2B**).

These results validate two fundamental requirements for strategic decision support: (1) **accuracy** approaching expert-level assessment, and (2) **consistency**, so that the query consistently generates identical assessments, eliminating the randomness observed in outputs from off-the-shelf LLMs. Together, these findings demonstrate that our augmented intelligence architecture delivers the reliability and reproducibility required to support and optimize real-world biopharma strategic decision-making.

Performance benchmark summary

Our augmented intelligence system achieved 89-95% accuracy and 83-94% reproducibility across three critical in-licensing dimensions in a real-world validation dataset. This represents a 2.5 to 4-fold improvement over baseline foundation model performance (23-36% accuracy).

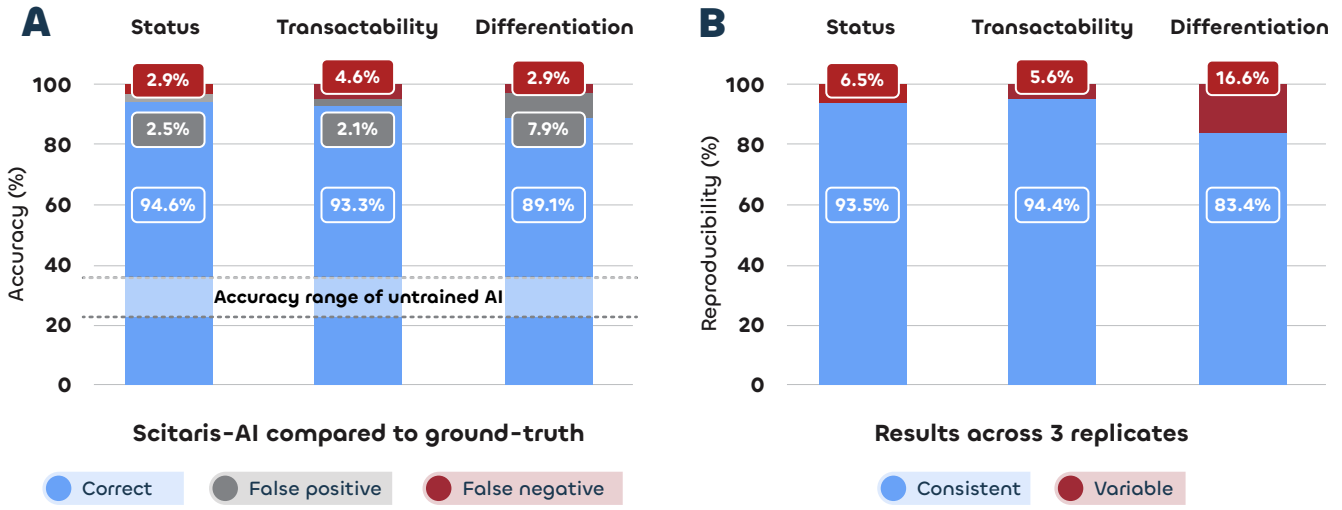

Figure 2. Quantitative results of accuracy and reproducibility of *Scitaris-AI* assessment of a real-world database.
A. Stacked bar charts showing *Scitaris-AI* accuracy compared to expert ground-truth per dimension (across three independent runs combined). Blue indicates correct classifications; gray indicates false positives; red indicates false negatives.
B. Stacked bar charts showing reproducibility across three replicates. Blue indicates assets receiving consistent classifications across all runs; red indicates variable results.

4. DISCUSSION: Implementation, impact, and outlook

The benchmarking results presented here validate that augmented intelligence architectures can deliver decision-support capabilities approaching expert-level performance. **With 89-95% accuracy and up to 94% reproducibility, *Scitaris-AI* addresses the fundamental limitations we identified in off-the-shelf LLMs, transforming unreliable outputs into consistent, actionable assessments.**

Scitaris-AI currently integrates our established workflow as a new layer in the expert-driven process our clients rely on. The three dimensions presented here – development status, transactability, and differentiation – showcase a simple proof-of-concept implementation. Our **modular architecture**, where each evaluative dimension operates with its own dedicated dataset and analysis protocol, enables systematic expansion of scope and complexity. The platform is already being used to support other strategic applications beyond in-licensing evaluation where structured, reproducible analysis adds value. Frameworks for additional decision dimensions are in place, and we are currently implementing weighted outcome scoring to enable more granular ranking of opportunities – **evolving the system toward a comprehensive prioritization tool.**

As the platform matures, our clients will benefit from continuously enhanced analytical capabilities, including broader and deeper analysis. PhD-level consultants remain central to the workflow, providing relevant inputs, interpreting outputs in strategic context, handling nuanced edge cases, and applying the judgment essential for final recommendations. Crucially, *Scitaris-AI* can be configured to reflect specific strategic priorities (therapeutic focus, geographic preferences, deal structure constraints, and portfolio gaps), ensuring alignment with each client's unique objectives. ***Scitaris-AI* ultimately serves as a flexible companion system that elevates the quality, breadth, and depth of expert analysis, supporting the rigor and consistency that confident high-stakes strategic decisions demand.**

Looking forward

The capabilities shown here are the foundation of *Scitaris-AI*. Beyond these proof-of-concept decision-making dimensions, *Scitaris-AI* already encompasses additional analytical capabilities that extend across the strategic landscape of biopharma business development. As we validate and deploy these modules, our commitment remains constant:
AI should enhance and standardize expert analysis while preserving the human judgment essential for high-stakes decision-making.

Conclusion

While off-the-shelf LLMs show 23-36% accuracy when judging in-licensing opportunities, ***Scitaris-AI* achieves 89-95% accuracy through systematic integration of curated data, expert frameworks, and structured analytical workflows.** This augmented intelligence approach elevates PhD-level expert assessment, enabling broader and deeper analysis while maintaining consistency.

Contact Us

About the Authors

Catarina Martins Costa, PhD
Consultant

Catarina earned her M.Sc. in Molecular Biotechnology from the University of Porto, having developed her Master's Thesis at the University of Edinburgh. During her PhD at the Vienna BioCenter, she used human neural organoids to study rare neurodevelopmental disorders. For her PhD work, Catarina was awarded the Vienna BioCenter PhD Award, the Impact Award of the City of Vienna, and the Austria State Prize.

Wignand Mühlhäuser, PhD
Engagement Manager

Wignand earned his B.Sc. in Applied Biology at Heinrich-Heine-University Düsseldorf, where he explored RNAi mechanisms and developed a strong interest in translational research. He completed his Master's at Albert-Ludwigs University Freiburg, focusing on synthetic biology and building his first optogenetic tools. During his PhD, he optimized and expanded optogenetic toolkits for controlling oncogenic signaling proteins and received the GBM Young Innovators Award.

The system described in this article was supported within the framework of the project "KI-Servicezentrum Berlin-Brandenburg" (funding code 16IS22092), funded by the Federal Ministry of Research, Technology and Space.

Responsibility for the content of this publication lies with the author.

With funding from the:

Federal Ministry of Research, Technology and Space

References

1. Mühlhäuser W, Vonnemann J. *Can LLMs judge in-licensing opportunities? Benchmarking prompt-engineered foundation models for biopharma strategy* (2025)

2. Berteletti, E. et al.. *McKinsey Talks Operation Blog: Breakthroughs in AI-augmented R&D: Recap from the 2025 R&D Leaders Forum* (2025)

3. Noble, J. *What is LLM temperature?* (2025)

SCITARIS
BIOPHARMA STRATEGY