WHITE PAPER

# RECURSIVE AI LEARNING MODEL FOR PRIMARY AND PROPRIETARY HEALTHCARE DATASETS

**April 1, 2025**

RegenMed

# TABLE OF CONTENTS

## Circles Current State

The patented Circles technical platform is a closed system enabling the collection, aggregation, and analysis of "real world" healthcare data from verifiable and independent primary sources. [1] Those primary sources generally include patients and their physicians in the context of a specific clinical intervention. Other possible primary sources include laboratory technicians and caregivers.

All such data is collected in the context of an Observational Protocol (OP). Each OP comprises several Surveys, [2] and each Survey comprises a number of "canonical" Questions. Each Answer to each Question is a Datapoint.

Each Observational Protocol is defined by a single anatomical region, a single pathology, a single treatment protocol, and one or two standardized outcomes measurements. The data collected against each OP in the context of a single patient is called a Case. Each Case is by definition longitudinal – outcomes data correlated to the clinical intervention and corresponding clinical hypothesis is collected over a year or longer. A Case will therefore typically represent 500 or more Datapoints, each inheriting the attributes of the OP and therefore well correlated to a specific pathology, clinical hypothesis, treatment, and patient cohort. [3] When a number of collaborating physicians collect data under the same OP, a statistically significant and clinically significant Circles dataset thus results. [4]

Any Circles dataset can be analyzed to support clinical decision-making, reimbursement, safety and efficacy claims, regulatory submissions, value-based care, social determinants of health, content for articles and conference presentations, training of AI healthcare models, etc.

Because all Circles datasets are originated and maintained exclusively within a closed and secure technical system, they are fully verifiable, transparent and proprietary. There is no need to import other data from any other source to make them clinically and statistically relevant.

## Adding Recursive AI Learning and Reasoning For Continual Improvement

### *General*

As AI learning and reasoning models continue to evolve, their application to Circles datasets offer several advantages. These include:

- Improved or new clinical hypotheses and Observational Protocols.
- Verifiable, hallucination-free evidence-based clinical decision support.

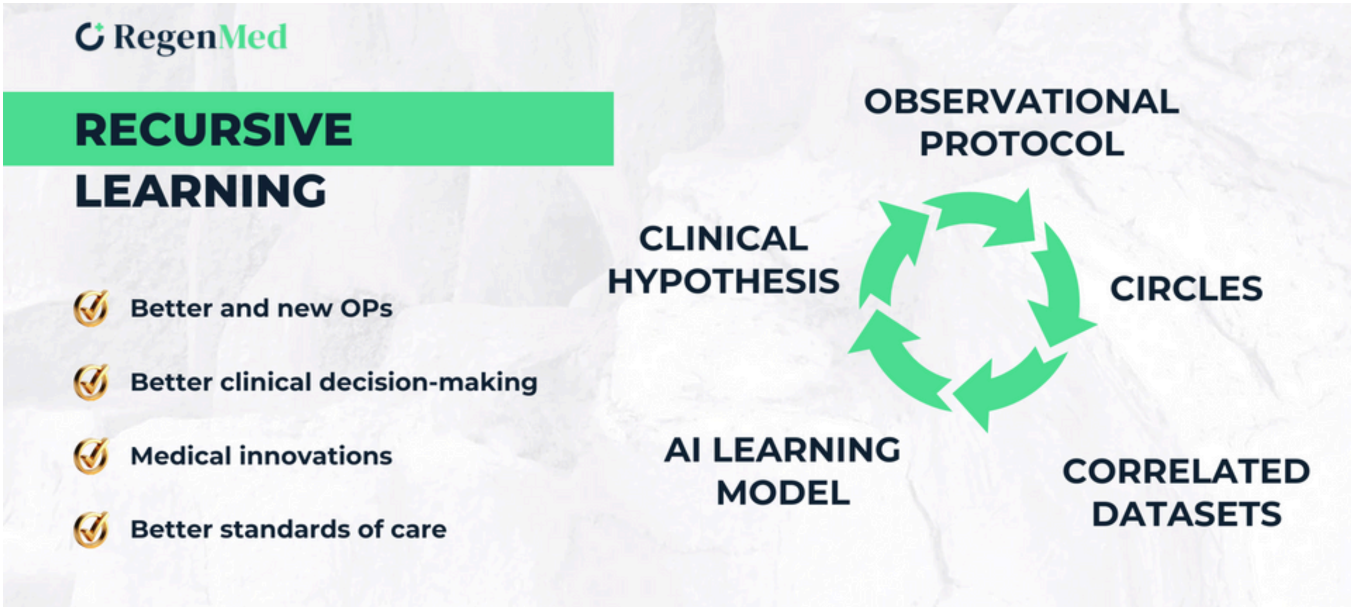
- Suggested improvements in existing medical products, different dosage regimes, new products, new diagnostics.
- Product-agnostic evidence-based support for value-based care.
- Support for regulatory submissions, post-market surveillance, product marketing claims, etc.

***Continual and Automatic Deepening of Clinical, Scientific, and Financial Value***

A defining feature of Circles datasets is their automatic enrichment through the continual accumulation of longitudinal and verifiable data. Physicians tend to see the same types of patients, and perform similar treatments. In addition, longitudinal outcomes data continues to be captured for clinical interventions already performed. Thus, Circles datasets inherently become more clinically and statistically significant with time.

The insertion into the Circles process of an appropriately structured AI learning/reasoning model will only enhance their power. A typical use case would be the ingestion of one or more related Circles datasets – for example, all datasets relating to total knee arthroplasty – by the AI learning/reasoning model to generate new clinical hypotheses, value based medicine decisions, cost versus outcomes analysis, etc. in the context of a specific knee pathology and treatment.

Thus, the AI learning/reasoning model is designed to function as part of a recursive, continually improving process. Proprietary, primary, and clinically relevant data is collected through everyday clinical interventions and stored on the Circles platform. This data is then analyzed by the AI learning/reasoning model, which identifies opportunities to improve Observational Protocols, as well as the quality and cost of patient care.



AI-generated recommendations are fed back into the system to inform the next round of data collection. As more data is collected and analyzed, the AI model continually refines its insights, providing increasingly effective and precise recommendations. This recursive process allows the model to learn from its previous iterations, enhancing both short-term decision-making and long-term outcomes analysis.

# THE CHALLENGES OF CURRENT “BIG DATA” APPROACHES

Much of the current $60+ billion (21% CAGR) healthcare data analytics market is fundamentally flawed due to two major factors: (i) although undoubtedly “big”, the underlying datasets are unreliable; (ii) ownership of those original datasets is disputed or disputable.

**Unverifiable and Incomplete Original Data + Undisclosed Algorithms ≠ Useful Data**

“Garbage in, garbage out” is as relevant to healthcare data as to any other form of processed data. However, the consequences in healthcare are of course serious. The large majority of healthcare data today relies on an amalgam of disparate non-verified (and unverifiable) original data sources. Moreover, those datasets have been “cleaned”, “synthesized”, “processed”, and otherwise manipulated through a variety of undisclosed algorithms. The resulting “big” datasets are licensed as clinically relevant. In reality, however, they cannot be due to serious deficiencies in the original data: material gaps, lack of meaningful clinical context, limited or no outcomes data (let alone outcomes correlated to the original clinical intervention), non-verifiability, and explicit or implicit bias.

Moreover, “big data” systems are often trained on datasets which may not generalize well to broader patient populations. Their recommendations can be difficult to explain, which undermines trust.

Machine learning and AI applied to “big data” does a good job of identifying statistical patterns. But those patterns follow Mark Twain’s well-known aphorism [5] in terms of clinical utility. Establishing testable causal correlations in healthcare requires the type of structured and validatable data available only from transparent and properly conducted clinical studies. The only valid test of any asserted causal correlation is the ability reliably to repeat it. The only useful environment for attempting those asserted correlations is in the clinic.

# THE CHALLENGES OF CURRENT "BIG DATA" APPROACHES



### Undisputed Ownership

The ingestion by AI models of large amounts of data has brought to the fore the critical issue of data ownership. [6] As a result, EMR vendors, hospital systems, payers, data aggregators, medical societies, registry sponsors, and numerous other parties will claim to own "their" data, and seek to monetize it on the basis of that ownership. Litigation and increased expense for all parties seeking to use such healthcare data will be the inevitable result.

However, at the heart of all important healthcare data is patient data and the data reflecting their physicians' clinical judgments. It is likely that the majority of patient data being used in "big data" is not properly anonymized, and/or is not associated with an adequate informed patient consent allowing the monetization of their data. Legislation and courts will likely come to recognize claims by patients on the value of their personal healthcare data even if anonymized. [7] Indeed, as that data becomes more personalized – genomics, proteomics, microbiomics and other "omics" – it will be impossible to assert that it has been anonymized. [8]

# BENEFITS OF CIRCLES RECURSIVE AI LEARNING / REASONING MODEL

### Data Verifiability and Integrity

An essential aspect of the Circle AI learning/reasoning model's reliability and utility will be the ability to trace all analyzed data back to its primary sources. Ensuring data verifiability provides several critical advantages:

- **Transparency:** By maintaining clear records of data provenance, stakeholders can confidently assess the validity of AI-generated insights.
- **Trustworthiness:** Clinicians and patients are more likely to trust and adopt AI-generated recommendations when they can trace those insights back to verified, primary sources.
- **Accountability:** Verifiable data establishes accountability, allowing clinicians to understand the underlying evidence supporting AI-driven recommendations.

# BENEFITS OF CIRCLES RECURSIVE AI LEARNING / REASONING MODEL



- **Regulatory Compliance:** Traceable data sources facilitate adherence to data governance standards and regulations, further enhancing credibility.

By definition, Circles employ rigorous data collection processes ensuring that every datapoint is verifiably linked to its original source. This commitment to verifiability enhances the overall robustness and reliability of a Circles AI learning/reasoning model.

## Data Ownership

As described above, Circles is a closed system; it does not import or ingest data from any other source. The physician's clinical data is entered by them in a fully validatable manner, and reflects their independent clinical/scientific hypothesis and judgment. A separate and informed patient consent is collected from each patient, and associated with each Case.

In addition, as discussed in a separate white paper, a blockchain/tokenization structure is being explored to fairly compensate and incentivize patients for the completion of outcomes surveys. This structure will reflect legal and ethical standards, with a focus on informed, voluntary, and revocable consent to data usage.

## Technical Elements

Technical aspects of the Circles AI learning/reasoning model will include:

***Data Quality and Validation***

Implement processes for validating and standardizing all incoming Circles datasets to ensure accuracy, consistency, and completeness. Utilize automated data curation systems that cross-reference input data with established medical guidelines, peer-reviewed literature, and other validated clinical datasets. Employ anomaly detection algorithms to identify and rectify inconsistencies or errors in the data.

***Causal Inference Modeling***

Develop AI models specifically designed for causal inference rather than mere correlation detection. Integrate causal reasoning algorithms such as Structural Causal Models (SCMs), Bayesian networks, and counterfactual inference techniques. Enhance recursive learning models by incorporating causal inference feedback loops, where causal relationships are continuously validated and refined over multiple iterations of data collection and analysis.

# BENEFITS OF CIRCLES RECURSIVE AI LEARNING / REASONING MODEL



***Enhancing Explainability and Trust***

Develop AI models with built-in explainability features to improve clinician and patient trust in the system's recommendations. Make model outputs interpretable and actionable for healthcare professionals. Use a combination of interpretable models (e.g., decision trees, rule-based systems) and post-hoc explainability methods (e.g., LIME, SHAP, attention mechanisms, and saliency maps). Provide clinicians and patients with comprehensive, transparent reports detailing how recommendations are derived and validated.

***Continual Learning and Feedback Integration***

Ensure that the AI model continually learns from new data and integrates feedback from clinical experts to enhance precision, relevance, and robustness of recommendations. Implement reinforcement learning and active learning frameworks. Develop systems for dynamically updating AI models based on clinician feedback, patient outcomes, and newly published research.

***Integrating Patient and Physician Ownership Rights***

Ensure that both patients and physicians have auditable ownership rights over contributed data. Provide a clear framework for sharing the value generated by the system. Design smart contracts using blockchain technology to automatically distribute benefits generated by AI-driven insights to patients and physicians. Provide transparent mechanisms for tracking data usage and resulting compensation where due. [2]

# ILLUSTRATIVE USE CASES

The AI learning/reasoning model can be utilized for many scientifically, clinically and financially important use cases. These include:

***Improved and New Study Protocols***

Suggesting new, and improving existing, Observational Protocols for any pathology, treatment protocol, or health/wellness goal. Recommendations can be adapted to specific patient demographics, co-morbidities, and treatment histories. Enhanced protocol design for clinical trials.

***Clinical Decision Support***

Evidence-based guidance delivered at the point of care to enhance standards of care. Recommendations can be integrated into clinical workflows.

***New or Improved Standards of Care***

Unbiased, longitudinal and transparent evidence for any pathology or health/wellness intervention, including complementary and alternative medicine. [10]

***Genuine Value-Based Care***

Compare products, costs, and outcomes for specific pathologies, treatment protocols and patient cohorts. This is far more meaningful value based care than, for example, mere readmission rates.

***Combined Primary and Adjunct Therapies***

Generate evidence-based treatment protocols correlated to long-term outcomes for combination therapies. [11]

***Medical Innovation***

Identify emerging patterns and therapeutic opportunities, facilitating the development of novel treatments, interventions, and diagnostics.

***Second Opinions***

Verifiable evidence-based second opinions for physicians or patients.

## Footnotes

1. Circles real world data is in sharp contrast to other healthcare datasets often referred to as "real world". The latter are an amalgam of prescription and other medical codes, registry information, and insurance claims to which various machine learning algorithms have been applied to approximate coherent datasets. However, these datasets are rife with gaps, errors and unverifiable (therefore untrustworthy) data. Moreover, the ownership of the data used in these datasets is often disputed.
2. Surveys are designed so that each one is to be answered by one person. For example, there are patient-facing Surveys and physician-facing Surveys. There may be a separate Survey for a laboratory technician or caregiver.Patient Surveys are typically of two types - Demographic and Outcomes. The same Outcome Survey is used to collect pre-procedure baseline information, and then that same information one, three, six, nine and twelve months post procedure. (It is possible to change the rhythms and duration of the Outcome Surveys.)
3. Other attributes assigned to each Observational Protocol include CPT, ICD, SNOMED, HCPCS and other medical codes. These attributes can be assigned at the creation of the OP or retroactively. This is useful for codes which differ from country to country.
4. Existing Circles technical functionality enables the burden-free collaboration among physicians sharing a common Observational Protocol.
5. "Lies, damn lies, and statistics."
6. See https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html, https://www.wired.com/story/thomson-reuters-ai-copyright-lawsuit
7. See for example https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-022-00848, and https://fortune.com/2025/03/28/23andme-bankruptcy-chapter-11-genetic-data-medical-ancestry-dna-silicon-valley-delete/.
8. See for example https://news.bloomberglaw.com/privacy-and-data-security/patients-advance-meta-lawsuit-over-collecting-health-information, https://www.communicationslitigationtoday.com/article/2024/06/14/summit-health-violated-web-users-privacy-by-monetizing-their-pii-class-action-2406130030, https://natlawreview.com/article/goodrx-agrees-pay-25-million-settlement-privacy-violations, and https://www.classaction.org/news/3.8m-davita-settlement-resolves-class-action-lawsuit-over-alleged-data-sharing-violations.
9. See separate RegenMed White Paper on Circles Blockchain/Tokenization.
10. Many, if not, most current standards of care have poor levels of evidence. Moreover, "Complementary and Alternative Medicine" has virtually no supporting evidence. (That does not mean that CAM therapies are not safe and efficacious; it only means that there is little evidence for them.) Meanwhile, forty percent of Americans use at least one CAM therapy. Fifty percent of physicians say they have recommended at least one CAM therapy to their patients during the prior 12 months.
11. As but one of hundreds of examples, total hip arthroplasty with a specified rehabilitation protocol.