AI Managing Agent-Based Healthcare Processes

Simon Grange[1], **Pearl Ryuawa**[2], **Safa AlAmeri**[1,3], **Rami Bahsoon**[1]

[1] University of Birmingham, UK
[2] New York University, Abu Dhabi, UAE
[3] Department of Health, Abu Dhabi, UAE

Abstract

This paper describes a methodology for evolving systems supporting governance, regulation, control, safety, and security in personalized healthcare. The principles of the 5 'Ps' in healthcare is a patient-centred approach encompassing predictive, preventive, participatory, personalized and psychocognitive care channels, focusing on individual needs and preferences. To embrace AI in a critical system, any stochastic advantage of time or resource optimization needs to be trusted, and this in turn needs deterministic consolidation – i.e. verification processes, which both secure the foundation of any novel system through offering reassurances of the reliance upon models and validation of those models in practice, since they may "drift" over time. Deviating from this original foundation can lead to errors, which need to be addressed for such a system to remain useful, indeed credible.

This review explores of how such governance becomes integral to developing new principles for responsible AI, inspired by personalized healthcare's 5Ps, which can be adopted for the managing of a common yet critical care path. This leads to many questions around the strategic, operational and tactical approaches, which are answered through providing a use case of dealing with a medical emergency to exemplify future approaches to agent-based healthcare management.

Introduction

Artificial Intelligence (AI) has revolutionized healthcare by enabling agent-based systems that optimize clinical workflows, enhance patient care, and drive medical research. However, the integration of AI into healthcare necessitates robust governance frameworks to ensure ethical deployment, regulatory compliance, and system safety. This needs to reflect a level of fidelity that matches or exceeds the level of clinical tolerances. It needs to display consistency and where possible modularity around standardization that assumes interuptibility.

This outlines a methodology for developing evolving systems that govern AI-driven healthcare processes using individual agents, ensembles, and governed sectors. It further explores ethical models and frameworks to ensure safety and consistency, with an emphasis on near-real-time feedback loops, managed by Intelligent Digital Twin(IDT) modelling.

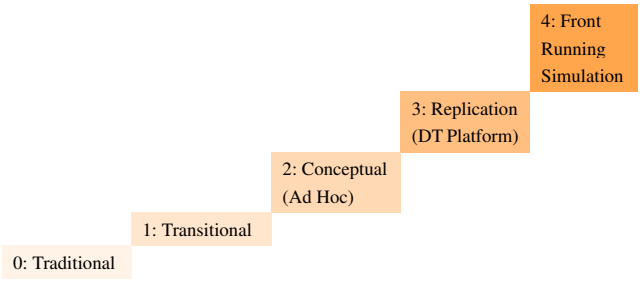

Table 1: Digital Twin Evolution – Physical to virtual maturity

These are evolving and as with any emergent technology, it is best rationalized by a roadmap to assist navigation both in terms of understanding the progress to date and its future trajectory.

At the top level, a project structure can encompass any scale of challenge and breaking this down to operational steps with goals, is symbolic of "team" effort, often managed by ensembles of agents, especially in the form of networks or clusters. These in turn need to manage discrete tasks, which can be performed by specific agents, either as complicated or simple tasks. The key with composite tasks structuring goals, is that there is the opportunity for emergent behaviour, suggesting a dimensional manifold. Through this, digital twins evolve, from traditional approaches to front running models. Table 1 above emphasises the progressive maturation of the digital twins.

Agent-Based Healthcare Systems

With the mantra of personalized healthcare, the purpose is to have agents working on behalf of the patient. Agent based self.X, that perform specific tasks, such as diagnosing diseases, managing patient records, or optimizing hospital workflows. These agents can operate individually or collaboratively as ensembles to address complex healthcare challenges.

Many attempts have been made to capture the processes (top down) e.g. technology & healthcare provider companies, compared with a bottom-up approach which focuses on THE patient.

Governance Challenges in AI

One sound argument for this is security. The rapid evolution of AI technologies introduces risks such as algorithmic bias, data privacy breaches, and paradoxically lack of transparency.

Governance frameworks must address these risks whilst fostering innovation and ensuring equitable access to AI-driven healthcare solutions (Tilala et al. 2024) (Ethical Considerations of AI — What Purpose Do Fairness Measures Serve in AI?, 2024). It must align and integrate both agent & human intent to protect individual identity e.g. GDPR and or confidentiality in accordance with 'know your client' (KYC) principles.

Digital Twin Technologies

Digital twins are virtual replicas of physical entities or systems that enable near real-time monitoring and simulation. In healthcare, digital twins can model patients, clinical workflows, or entire hospital systems to provide actionable insights and improve decision-making processes (Naik et al. 2022). They may benefit the patients, care providers, insurers and health authorities, responsible for managing demand.

As seen in Table 1 above, this is not a new concept, but an evolving one. as it is primarily commercially motivated, though safety is also very important. This has been driven by the development of machinery such as aircraft engine monitoring technologies, practically implementing Internet of Things (IoT). Below, this evolving methodology is expanded upon to justify the increasing complexity of such systems and the associated protection mechanisms these can offer.

Operational Methodology

The operational 'middle ground' between such grand strategy and practical tactical decision-making needs to also support the 'System Architecture for Governance'. In time this will focus on agents managing algorithms though the future of intelligent digital twins (IDT) is susceptible to the age-old issue of data quality. This needs to be addressed first and foremost with data quality assurance and structures such as those outlined in figure 1, which will play a key role in future. It is upon this foundation of validated data that the agents can get to work building and testing hypotheses – modelling actual and then, potential outcomes. Aligned with those agents, is the hierarchical governance system which figure 1's framework details. Simply put the agent must first develop autonomous competence with oversight, described by the basic control unit (levels 1 – 3), before it can effectively work as part of an ensemble which work through levels 4 – 6 to ensure integration and governance of the ensemble. Finally, levels 7 -9 represent harmonisation with the ecosystem in its entirety.

Individual Agents

Agents are designed as 'narrow AIs' with specific 'types'. They are thus associated with specific "payloads" i.e. rights of access to workflow components, such as databases and processes to fulfil certain roles e.g., diagnostic tools or patient monitoring systems.

Each agent is therefore able to be equipped with machine learning algorithms as part of the payload tailored to its function. Agents must however comply with data privacy regulations, like GDPR and HIPAA (Ethical Considerations of AI — What Purpose Do Fairness Measures Serve in AI?, 2024) and this is both monitored by the governance agents (level 3), the ensemble agent governance (level 6) and strategically embedded in the system (level 9) in accordance with the capabilities.

Ensembles

Collaborative functionality as achieved through the ensembles, consisting of multiple agents working together to achieve complex goals such as personalised treatment plans. These may be simple, complicated, complex or chaotic as defined below.

To coordinate these, it is necessary to define the meta structure and need for coordinating agents. Inter-Agent Communication via secure protocols ensures seamless communication between agents, while maintaining data integrity (Cloud Security Alliance 2025).

Data Governance Compliance Framework

This should not differ greatly from that required for the rest of society and evolving infrastructure to prevent unnecessary costs. Starting with data management ('garbage and garbage out'), a data mesh, which ensures that the operational data can be managed through micro services, provides control over the analytical processes for the analytical data which, once raw data is cleaned, should adequately allow both events and entities to be recognized within specific domains.

The use of Kafka™ is a very popular approach, providing open-source management of events in near real time. Such activities can be enhanced with the use of the AI tools and the enabling team, which will likely be Multidisciplinary.

The systems in healthcare involve autonomous entities (agents), usually represented by software classes.

They may consist of consulting advisers as well as those exercising 'best practice' under the supervision of a Federated Governance team, managing the domain bounded context.

Framework Features

Extending across such a wide area as the domain of healthcare requires a consistent approach. The framework is therefore integrating data governance and its adoption for responsible data management as described above, whilst ensuring data and privacy protection. This aligns well with the ethical AI development, but it is only effective if complied with.

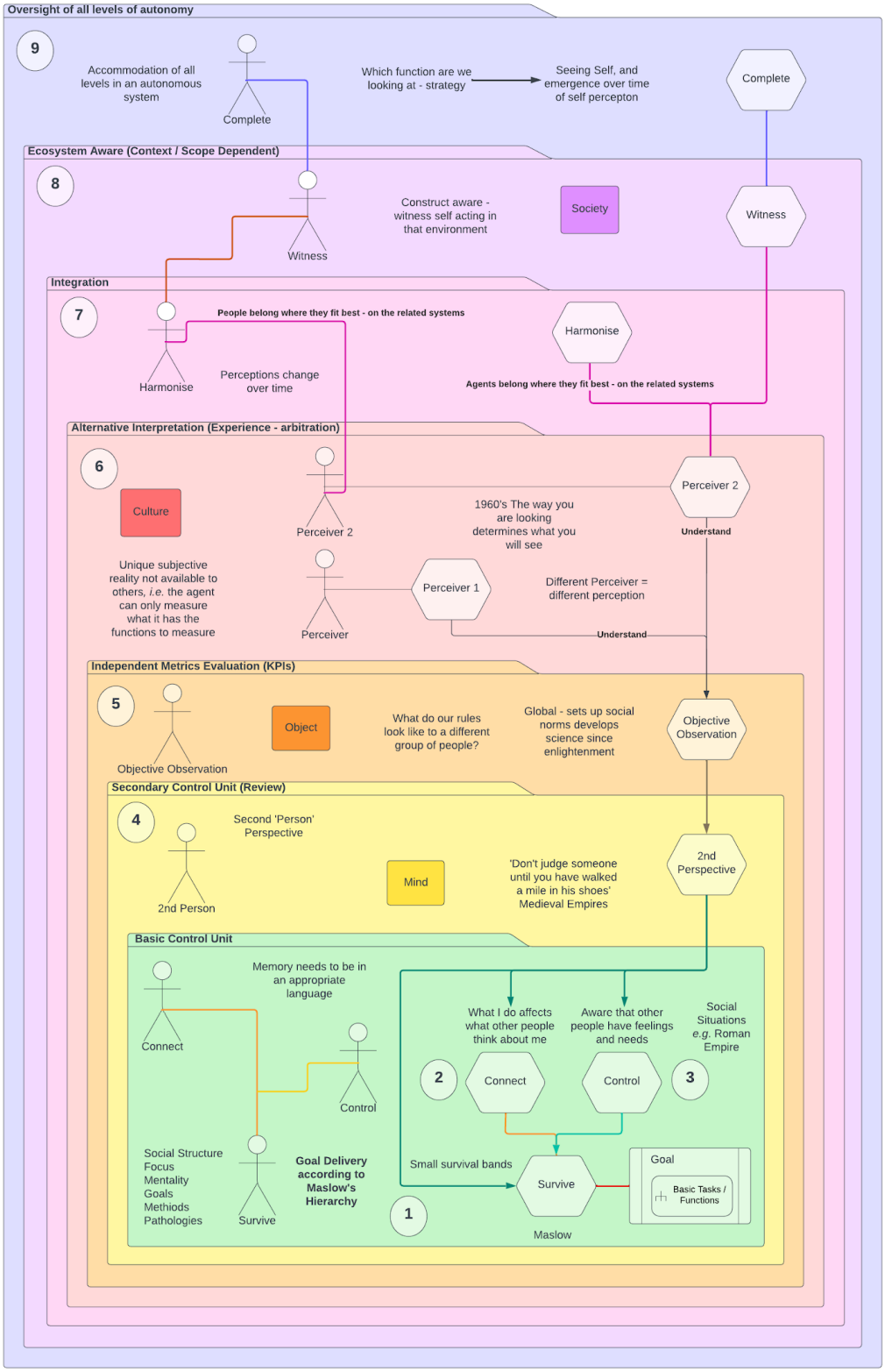

Figure 1: Data Governance Compliance Framework

So far, the process for this has been generated by regulatory authorities in accordance with the wide recognized general data protection regulations (GDPR), which is managed across Europe and elsewhere as well as the California Consumer Privacy Act (CCPA), which carries favour in the United States. These standards for privacy explicitly relate to consent, data accuracy, limitations of storage and the issues of non-compliance. The significant difference is that the European model is one of "opt in" whereas the American model is one of "opt out".

This aligns more with their industry perspectives. In addition to this, the role of the EUAI Act is more specific to the potential harm that AI could cause based upon their risk.

Therefore, any modern system implements both software and hardware in accordance with such principles to ensure the AI driven data governance remains compliant for automated data quality management and AI enhanced data lineage tracking. The AI is used to support regulatory compliance monitoring.

Governed Sectors

Because of the composite nature of the ensembles of agents, their collaboration for integration of Sector- Based Regulation of healthcare processes is itself divided into sectors (e.g., diagnostics, surgery, administration), each governed by specific AI regulations for goal achievement to reflect the purpose of a sector. This leads to Oversight Mechanisms.

Ethical review boards oversee sector- specific AI applications to ensure compliance with established standards (Cajueiro and Celestino 2024). It is therefore essential to develop oversight that is consistent and ensure that the Digital Twin (DT) maps onto the established processes of the Institutional Review Board (IRB).

Figure 4 outlines the three basic tiers at the level of the individual agent, its ensemble and as a global group, composing the governance hierarchy in terms of nine distinct levels.

Transparency

This tends to run counter to the conventional private sector practice, and so although transparency is critical for building trust in AI systems, developers must disclose their methodologies, data sources, and decision-making processes (Hat 2025; Lumenalta 2024). This has been a factor in the recent LLM "arms race", and LLMs are becoming an integral part of the communications toolkit in healthcare.

Strategies include 'open access' to algorithmic logic, which may be proprietary and so may require special protection and clear documentation of data handling practices which is expanded upon in the "data contracts" below.

Accountability

Accountability mechanisms ensure organizations take responsibility for AI outcomes. Based on the need for transparency, manifest as auditable algorithms and compulsory mechanism for reporting adverse events to oversight bodies (Request for Information to the Update of the National Artificial Intelligence Research and Development Strategic Plan, 2022). To this end the agent governance mechanism should mirror the established processes of responsible officers (RO), which has been established at the human level.

Bias Mitigation

Bias in AI potentially leads to inequitable healthcare outcomes. Addressing bias therefore involves diverse training datasets, to avoid "selection bias", using similar data sets to those in the real world. Regular audits of algorithmic fairness (Lumenalta 2024) can be conducted, and these may be used to identify deviation and drift with specific tools such as SHAP.

Digital Twin Integration

Digital twins, including the "Intel" payload of their agents, which enables near-real-time feedback loops by simulating healthcare processes dynamically. Key components include:

- **Feedback Loops:** Continuous monitoring through digital twins providing real-time insights that inform decision-making and system adjustments, for predictive medicine. The key is to standardize and automate through **robotic process automation (RPA)**.
- **Patient Modelling:** Digital Twins replicate individual personalised patient profiles, based on medical history, genetic data, and real-time health metrics e.g. heart rate (HR) and blood pressure (BP) etc, so personalized treatment plans are created which support proactive interventions, so mapping the user profile to prospective best available outcomes.
- **Workflow Optimisation:** Hospital workflows modelled as DT, identify bottlenecks and optimize resource allocation, playing apart in preventative solutions.

Implementation Strategies

The Governance processes must align with the healthcare system's expectations and so involve policy development. Establishing policies that define acceptable use cases for healthcare AI, and Stakeholder Collaboration aim for synergy through engaging developers, clinicians, ethicists, and patients in decision-making processes.

The Regulation Process involves Compliance Monitoring by implementing mechanisms to ensure adherence to regulations across all stages of development (Müller 2023). Ultimately this underpins standardisation, i.e. developing standards for data quality, algorithm transparency, and system interoperability.

Safety requires regular Risk Assessment; identifying potential risks associated with AI tools, such as errors in diagnosis or breaches of privacy and the use of Testing Protocols to conduct rigorous testing under simulated real-world conditions before deployment (TRL5) (Müller 2023).

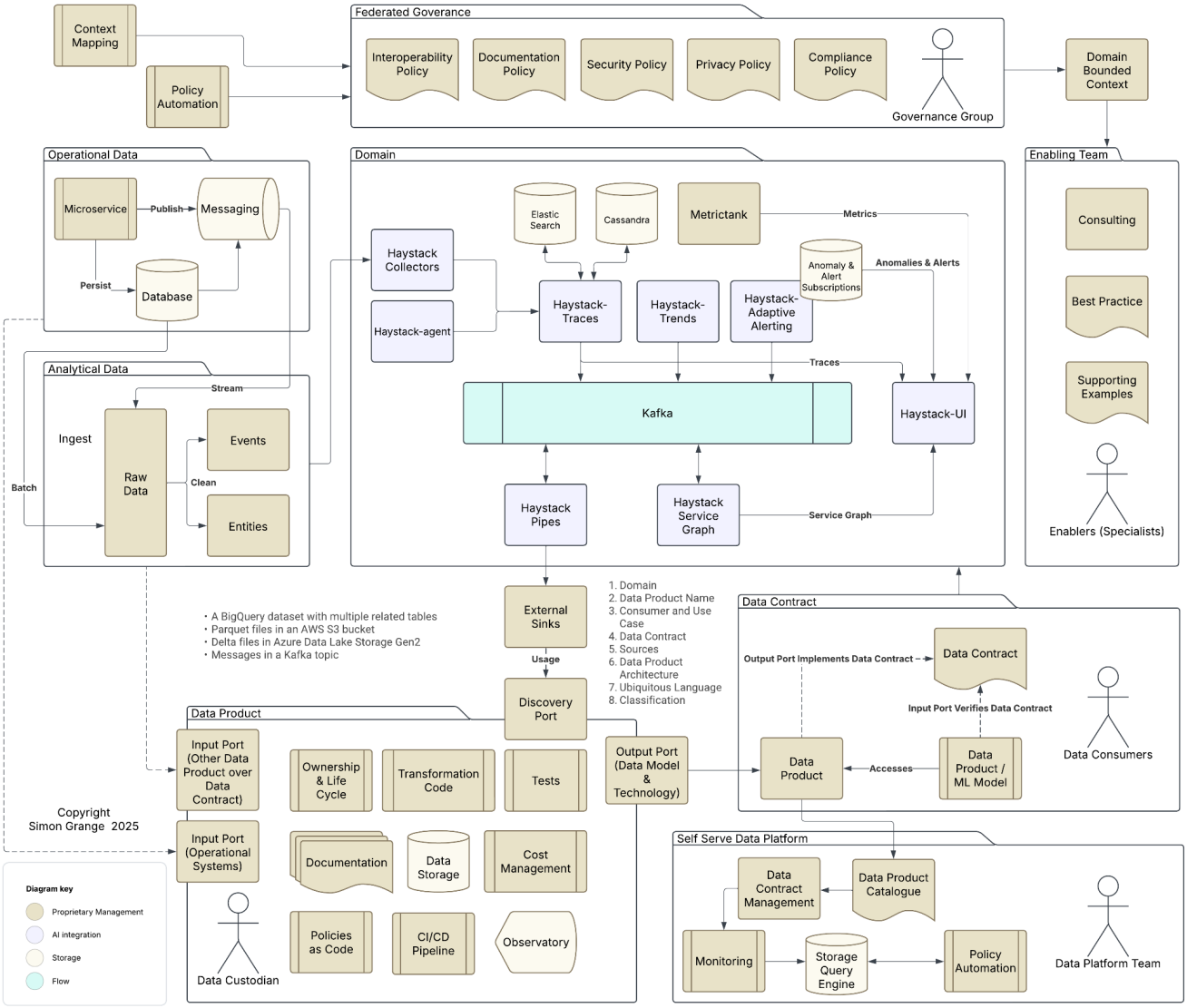

Figure 2: Data Governance Compliance Framework

Finally, the security management depends on Data Protection by employing encryption techniques to safeguard sensitive patient information and broader Cybersecurity Measures to protect systems from malicious attacks that could compromise functionality or data integrity.

Stakeholder Collaboration

AI governance therefore requires collaboration among stakeholders including policymakers, healthcare providers, technologists, and patients reflecting the participatory nature of personalised healthcare (Request for Information to the Update of the National Artificial Intelligence Research and Development Strategic Plan, 2022). These strategies include interdisciplinary research initiatives and public consultations on ethical standards.

Standardized Benchmarks

Establishing universal benchmarks for safety and efficacy ensures consistency across AI applications in healthcare (Hat 2025). Such benchmarks should evaluate everything from algorithmic robustness to fairness across demographic groups. The best way to achieve fairness is to allow for Learning algorithms and models.

Ethical Oversight Boards

Institutional Review Boards (IRBs) and authorities, enforce standards set by these dedicated regulatory boards that oversee the development and deployment of AI systems in

healthcare sectors. Their responsibilities include reviewing ethical implications of new technologies and monitoring compliance with regulations (Naik et al. 2022). However much of the equipment and processes are managed at the national level. An ECG machine will provide a routine cardiac trace and offer an algorithm-based interpretation as guidance. When interpreting this, there is a 'Human-in-the-loop' (HITL) and they are legally responsible. The circumstances are very different, if a device is controlling the flow of a medication into the bloodstream, and so the medical devices agency will set stringent criteria for the parameters and acceptable tolerances.

Data Compliance thus includes various policies for security, privacy and compliance ensuring interoperability and clear documentation. Such an approach allows data to be managed in many different formats and with data contracts, ensuring its provenance. This assures that the "data product" meets adequate quality assurance and again has a data custodian who is responsible for this. Since this model can be applied universally, it can become an integral part of the intelligent digital twin (IDT) pathways and other process mapping, such as robotic process automation (RPA). Whilst RPA is a logical extension from conventional mandraulic processes, the transition to level 3 (Replication) DT still needs human interaction, before full level 4 automation.

Tactical Implementation - A Case Study

As a real-world case study, considering how an AI support tool would implement this; it will be aware that some solutions are simple, for example a trauma case presents to an emergency department with an uncertain combination of diagnoses, as they could have a range of potential injuries and time is a factor in their acute management.

The treatment strategy is not simple until broken down into basic sequential tasks. Developing strategies that are flexible enough to address simple, complicated, complex and chaotic patterns of events. In such a case presentation, the needs are seamlessly integrated and need to be teased apart for experts to address them separately. The basic Cynefin model (figure 4 below) describes the relationship of what class of event can be managed using single agents and which require ensembles of agents. The aim is to align with established protocols, such as Advanced Trauma Life Support (ATLS).

Community-Driven AI Models

Collaboratively developed models like this one emphasize transparency, while embedding safeguards against misuse, benefitting from years of human experience. For example, open-source frameworks for bias detection and public datasets anonymized for privacy protection (Hat 2025) should be routinely deployed. The main purpose is to emphasize the development of consistent systems and tasks tools that can be advanced through self-learning. The other aspect of this is the ability to break down challenging solutions such as chaotic problems to their complex components as defined by ensembles to then defer to complicated or simple task management by individual agents using dedicated "payloads". These are specifically designed for routine

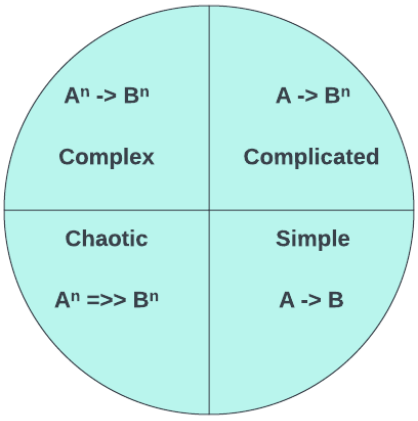

Figure 3: Cynefin Model

task management. The challenge is to be able to use an LLM prompt template to create suitable discourse to decipher the actual meaning of the challenge. This needs to refer to the discreet ontology to infer (Wills et al. 2004), and which can be updated (learning process).

The Cynefin model highlights that some problems cannot be reduced to simple or even complicated ones, managed by individual agents, but do require supervision of the ensembles for their integration, set within the context of their ecosystem that ultimately provides oversight of this. The purpose of this approach is to ensure autonomy with completeness. The original model recognizes the risk of disorder.

Digital twins have been used to simulate surgical procedures, enabling surgeons to rehearse complex operations virtually before performing them on patients, i.e. a simulation model (Grange97 1997). Two decades on, from "just-in- case" training to the "Just-in-time" copilot style of modelling as a surgical mentor, to deal with the transition from a conceptual DT to replication of the real-world real-time event modelling DT platform for training.

In such an initial incomplete clinical scenario, recommending potential 'safe' options is analogous to a collision avoidance mechanism in aircraft – the medical equivalent of "pull up!" to Front Running Simulation.

Digital Twin Applications in Surgery

Front Running Simulation (FRS), can provide a range of predictive scenarios, recommending likely response to possible actions, such as perceived rate of blood loss and the timing of surgery, based on an individual's physiological parameters and their rate of change.

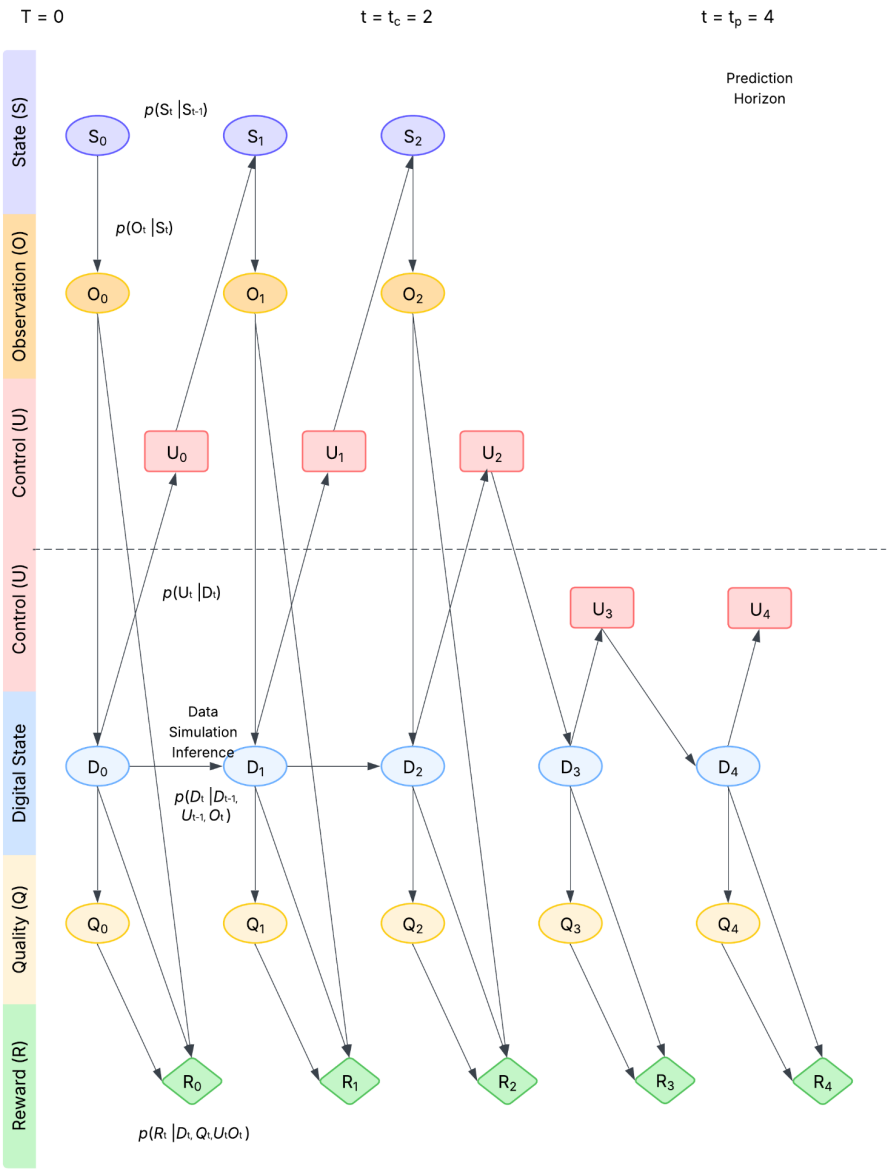

Figure 4: MAPE-K Cycles for Agent Payloads – demonstrating how each agent must "learn" and adapt.

The progressive building out of these approaches must use historical data in combination with patient specific data, since there may be 100,000 "concepts" that are known, even if only 10 are relevant at the time. This dimension reduction, in the information Science Integration for the Business Processes space is key to avoiding the "just too late" scenario of critical information being overlooked. Usually, the clinicians are aware of the key issues, but acute situations, errors of omission due to cognitive overload are a known risk, and as AI improves in the management of such care paths, not using IDT may be considered negligent in future.

To expand upon the information in figure 4 above, the value of this relates specifically to the ability of individual agents (and subsequently their ensembles) to learn over time. With any such system, the state can be changed through observation so that the Digital Twin's "digital state" will be updated along with the actual "control". This may also be reflected in terms of the change in data and recognition of the relevance "reward".

The coordination of this is therefore complicated at best, and potentially complex, where ensembles may display emergent behaviour. attention is therefore drawn in figure 5 to how such AI tools can be applied, so it becomes clear which areas will be affected and require monitoring.

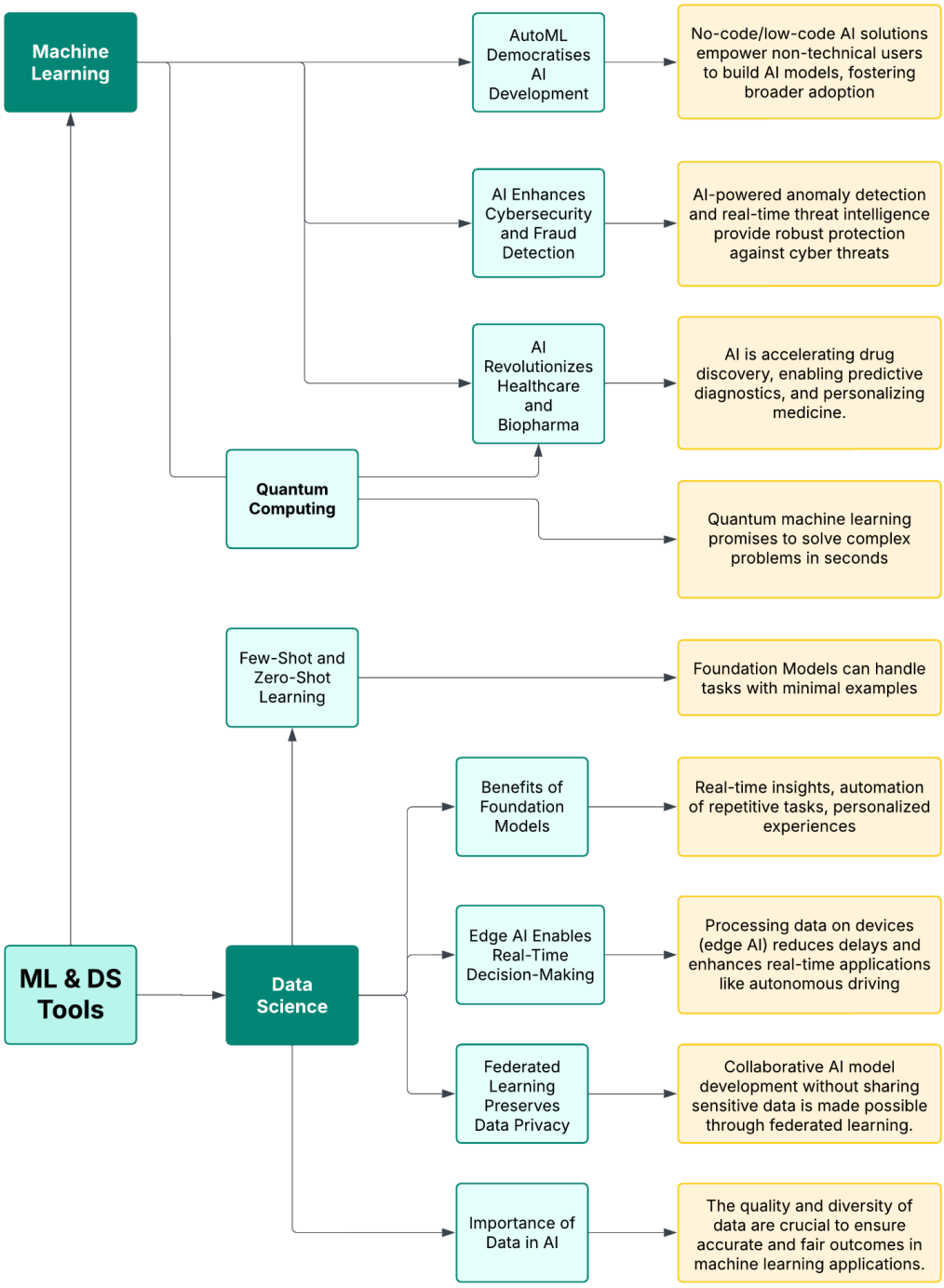

Figure 5: AI Tools using Machine Learning and Data

Broadly speaking, this has been broken down into Machine Learning (ML) and Data Science (DS) and the machine learning aspects of this are, except for quantum computing, which is still in practical infancy in terms of delivery, focusing on how AI can affect healthcare and biopharma through accelerating drug discovery, improving diagnostics and personalised medicine.

It is not possible to consider healthcare in splendid isolation. Rather it is also necessary to consider the use of AI for enhancing cybersecurity, with real time threat intelligence

and service protection against cyber threats, since even the UK National Health Service (NHS) has been a victim of this.

Finally, the concept of auto ML can democratize AI development. Having implemented such systems, the role of Data Science continues to change with great emphasis on 'zero shot learning' and the benefits of large foundation models, which are being developed from healthcare repositories.

Two other key factors in Data Science are relevant. The first being the increasing advancements of technology for 'Edge AI' enablement so near-real- time decision-making can be managed through data processing on Internet-of-Things (IoT) devices. Also taking advantage of the fundamental issue of privacy using federated learning, where collaborative AI models can be developed without sharing sensitive data.

Conclusion

The governance of agent-based healthcare processes powered by AI is a misnomer. There are many embedded AI systems performing unique and discrete roles, this leads to the concept of augmented intelligence. Ensuring data quality is fundamental to the success of any data science performed, and AI has its place in the management of this data irrespective of whether this is the healthcare environment or not.

A three-tier framework for governance that ensures agents can manage their own governance and yet align with the established models for responsible officers means that a degree of "agent self- awareness" is necessary for it to have insight into how well it is performing, against established benchmarks as part of progression towards a flexible dynamic FRS (level 4 automation) platform.

The next stage of this is to consider how ensembles can maintain their own governance, whilst supervising the basic control unit of individual agents, through reviewing with KPIs and if necessary, arbitration. This requires a comprehensive methodology that integrates ethical frameworks, safety measures, and is developed under the broader moniker of digital twin technology.

Fostering transparency, accountability, and collaboration among stakeholders, ensures the responsible deployment of AI systems where trust develops that prioritizes patient well-being while advancing medical innovation. Future research will focus on refining digital twin capabilities for broader applications in healthcare through defining standardization and interdisciplinary approaches to ethical AI governance.

References

Cajueiro, D. O.; and Celestino, V. R. R. 2024. A Comprehensive Review of Artificial Intelligence Regulation: Weighing Ethical Principles and Innovation. SSRN Scholarly Paper 4950727, Social Science Research Network.

Cloud Security Alliance. 2025. AI Safety vs. AI Security: Navigating the Differences — CSA. Retrieved May 14, 2025, from https://cloudsecurityalliance.org/blog/2024/03/19/ai-safety-vs-aisecurity-navigating-the-commonality-and-differences. Accessed May 14, 2025.

Grange97. 1997. Virtual reality, a training world for shoulder arthroscopy. Retrieved May 14, 2025, from https://personalpages.surrey.ac.uk/r.bowden/publications/vrsig97old/proceed/033/grange97.html. Accessed May 14, 2025.

Hat, H. S. . R. 2025. The Ethics of Open and Public AI: Balancing Transparency and Safety. Red Hat blog, January 28, 2025. Retrieved from https://www.redhat.com/en/blog/ethics-open-and-public-ai-balancing-transparency-and-safety. Accessed April 2025.

Lumenalta. 2024. Ethical Considerations of AI — What Purpose do Fairness Measures Serve in AI? https://lumenalta.com/insights/ethical-considerations-of-ai. Accessed November 22, 2024.

Müller, V. C. 2023. Ethics of Artificial Intelligence and Robotics. In Zalta, E. N.; and Nodelman, U., eds., *The Stanford Encyclopedia of Philosophy (Fall 2023 Edition)*. Metaphysics Research Lab, Stanford University.

Naik, N.; Hameed, B. M. Z.; Shetty, D. K.; Swain, D.; Shah, M.; Paul, R.; Aggarwal, K.; Ibrahim, S.; Patil, V.; Smriti, K.; Shetty, S.; Rai, B. P.; Chlosta, P.; and Somani, B. K. 2022. Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? *Frontiers in Surgery*, 9.

Tilala, M. H.; Chenchala, P. K.; Choppadandi, A.; Kaur, J.; Naguri, S.; Saoji, R.; and Devaguptapu, B. 2024. Ethical Considerations in the Use of Artificial Intelligence and Machine Learning in Health Care: A Comprehensive Review. *Cureus*, 16(6): e62443.

Wills, G.; Woukeu, A.; Bailey, C.; Ong, A. L. S.; Carr, L.; Conole, G.; Hall, W.; and Grange, S. 2004. Ontological Driven Learning Agreements. In *Proceedings of the Conference on Virtual Reality and Education*, 217–222.