Enabling scalable AI for Digital Health: interoperability, consent and ethics support

Zoran Milosevic
Best Practice Software, Australia

Abstract — **This paper proposes an approach for building scalable AI applications in digital health, with a specific focus on addressing interoperability, consent and ethics challenges. These challenges need to be considered in the context of increasingly available tooling for streamlined model development, training, validation, and deployment, while accommodating novel solutions for explainable AI support for clinicians. Such an approach is required because digital health ecosystems involve many data type created by different systems, and often used as part of workflows over different jurisdictional boundaries. Interoperability solutions are needed to support technical and business agreements between parties providing data and services, including knowledge intensive services, such as ML and AI. Computable expression of consent and ethics policies are needed to control how patient information is used, including compliance with regulative rules, possibly from different policy contexts. Our approach, based on the latest interoperability and enterprise policy standards may provide a useful guidance for the practitioners building scalable AI solutions for digital health.**

Keywords – interoperability; digital health; consent; ethics deontic logic; ODP enterprise language (ODP-EL); AI; FHIR.

I. INTRODUCTION

A. Problem

There is increasing level of tooling available to support building AI applications. This includes streamlined model development, training, validation, and deployment such as available by Microsoft Azure [1], Amazon [2] and Google [3]. AI applications in digital health however require addressing additional challenges. The first one *interoperability*, from technical, information (i.e. semantic) and business perspectives, needed to support aggregating and processing data from many sources and sharing them across applications and users in a digital health ecosystem. The second one is supporting consumers in expressing their *consent* preferences to specify finer level of control over how their personal health data can be shared or exchanged. The third one is supporting transparency about *ethics* principles, in the environment involving increasing level of automated decision making and multiple stakeholders.

B. Key contributions

This paper is focussing on these three factors as specific enablers for building scalable AI applications. The paper provides two main contributions:

- A set of architectural and implementation guidance in support of building interoperable and open digital health ecosystems, accommodating many different types of AI applications
- computable expressions of policies associated with consent and ethics requirements, so that such expressions can be integrated with the digital health enriched with analytics and AI applications

The paper addresses software modelling and implementation gaps in these areas. In particular, the paper addresses the lack of the expression of computable policies in support of consent and ethics policies that govern access to, and exchange of, personal health information. We have recently proposed an approach [12][16], based on the RM-ODP enterprise language standard [4] and in this paper, we further elaborate on this approach in the context of digital health applications for primary healthcare domain covering requirements of general practitioners. Such applications are often referred to as practice management software (PMS).

C. Background

The paper is in part motivated by our research and practitioner experience in addressing digital health problems over last 15+ years or so. This period has seen:

- the increasing focus on interoperability as a way of addressing the fragmentation of data and applications in support of healthcare, as evident by the prominence of several interoperability frameworks [19] and more recently by the emergence of new HL7 standard, Fast Health Interoperability Resources (FHIR) [5]
- many new technologies, e.g., cloud, streaming, IoT, distributed ledgers, digital twins and new generation of ML and AI, that have created fertile ground for
- stronger involvement of patients and consumers, in accessing, controlling and making decision about their data, as part of their healthcare or wellness, in response to new regulative rules and also COVID19 issues [41]
- various regulative and normative approaches such as the US 21st Century Cure Act [13] and the European GDPR Regulation [14].

These technological, consumer-centric and standards developments, along with the recognition for new models of care, including care coordination across providers, personalisation, and genomics, are leading to the emergence of new generation of digital health ecosystems. These ecosystems provide foundations for safer, more efficient, and patient-centred care, with AI applications allowing for further benefits in support of better care, e.g. clinical images analysis in diagnostic systems, detecting bone structures, identifying eye disease, sepsis prediction in ICUs etc.

D. Scalable AI

The specific AI applications in healthcare mentioned above address discrete use cases, each of which has its own set of data, specific computation models, validation criteria etc. In general, however, AI can be applied to multiple set of data, including real-time data, across different clinical systems and services in the ecosystem, and possibly according to different patient consent and privacy preferences as part of continuity of care. These applications

also need to respect and be compliant with appropriate ethics or legal requirements.

This requires a *scalable AI approach*, which refers to the ability of algorithms, data, models, and infrastructure to operate at the size, speed, and complexity required for the problem across the organisations or even jurisdictions [6].

There is an increasing number of solutions that can be facilitated to adopt and enable scalable AI within healthcare organisations [6][7][10], allowing for the streamlined model development, training, validation, and deployment. There is also active research in explainable AI [9] and we expect these solutions to increase trust among clinicians in adopting AI as an additional tool in their practice.

The scalable approaches to AI however require agreement about the semantics of health information to be used, which has been one of the main interoperability challenges in digital health over decades. There is also further requirement about business agreement related to interoperability which is needed to support seamless operations across organisational boundaries, including how to handle patient *consent* preferences across different healthcare providers. Further challenge, brought by big data, analytics and AI is how to 'bake in' *ethics* principles as part of AI based solutions.

There are currently many references focusing on the identification of key ethics principles, such as privacy data protection, fairness, contestability, compliance and accountability [17][21], but there is increasing recognition for the need to support mechanisms for translating these principles into computable expression, to allow developers to include these as part of their AI solutions, as also stated in our recent study [12][16].

E. Paper structure

This paper is structured as follows. Section II highlights the holistic view on interoperability as an enabler for the creation of sustainable digital health ecosystems, including the role of the HL7 Fast Health Interoperability Resources (FHIR®) standard. Section III describes the role of consent and describes an approach to a computable expression of policies in general, and consent in particular, which can add trust to users about controlled access to their data. Section IV provides an 'ethics by design' methodology that can be used in guiding practitioners in broader policy context associated with ethics, and specifically in translating ethics principles into computable expressions to ensure responsible use of AI. Section V provides discussion about how our approach can be applied in the context pf practice management systems, i.e. digital health system centred around primary care providers. Section VI discusses several additional challenges, and Section VII provides summary and outlines future work directions.

II. INTEROPERABILITY – FOR BUILDING SUSTAINABLE DIGITAL HEALTH ECOSYSTEMS

A. General considerations

Healthcare delivery involves many providers on patient journey, such as general practitioners, specialists, allied health and many support staff, all of whom are involved in access to and exchange of patient health and personal information, while respecting applicable set of clinical and administrative policies. This, coupled with the digital applications developed by different vendors and at different level of maturity presents a significant complexity for people and organisations to interoperate, in support of more effective, efficient, and safe delivery of care.

In order to address dealing with different aspects of interoperability, from technical and organisational perspectives, several interoperability frameworks were published to assist various stakeholders in the understanding of their interoperability concerns and guide them to their solutions, as summarised in [19]. In addition, several health standards were developed by HL7 international, most notably FHIR [5], which is now gaining significant adoption by wider developer community.

B. Interoperability Frameworks

In general, interoperability can be defined as [19]:

The continual ability of an organisation (or a system) to use or offer business (or technical) services from or to another organisation (or system) and accordingly, exchange information (or data) with other organisations (or systems) to achieve a specified purpose in a given context.

This definition caters for three different perspectives of interoperability, the organisational, information and technical interoperability. The organisational perspective is about specifying the business context, legal and policy issues of relevance for understanding, specifying and deploying digital health systems. This typically includes concepts for the expression of business processes, business services, business policies and organisational structures, applicable to the intra-organisational, inter-organisational and cross-jurisdictional interactions.

The information perspective is focused on the semantics of information used for representing clinical and administrative concepts, description of key information components and their relationships, e.g. medication, allergy intolerance, vital signs, appointment and so on. Typically, the information components will refer to certain artefacts in the organisational perspective, e.g. an information component referring to a referral or an appointment.

The technical perspective is concerned with developing applications and technical services that implement enterprise models defined as part of the organisational perspectives and handling information components defined from the information perspective. It is also about specifying technical infrastructure components such as cloud, mobile devices, IoT and so on, and conformance requirements.

It is to be noted that these are different, but related architecture abstractions that can be applied to any digital health ecosystem, to facilitate the separation of concerns among those interested in business, clinical or technical aspects when designing, building and deploying services in such an ecosystem.

C. Fast Health interoperability Resources (FHIR)

HL7 FHIR [5] is the latest HL7 standard which addresses information requirements in terms of the specification of information components, rereferred to as 'FHIR resources', and their relationships. FHIR is also dealing with technical interoperability owing to several infrastructure components proposed to support building solutions while leveraging the latest web standards and applying a tight focus on implementability, as discussed next.

The main modelling concept in FHIR is called a *resource*, which is an information component that can be used to exchange and/or store healthcare data. Resources cover clinical and administrative components but also include several foundational components needed for the overall infrastructure of the FHIR specification [5]. The mainstream resources are focussed on clinical content models, but it is increasingly recognised that many digital health applications require explicit support for defining and implementing services and processes (workflows), with additional support for the expression of enterprise policies that apply to those involved in delivery of healthcare, including patients themselves. FHIR provides limited support for such policy concepts at present but it is excepted to evolve, as a result of its use and community feedback.

Many digital health projects are now looking at using FHIR for new digital health applications, while leveraging legacy systems, and the standard is gaining significant interest and adoption. In fact, FHIR is becoming foundation for many digital health ecosystems.

One example is Health Concourse [7], developed for the purpose of linking and processing data coming from many data sources, aggregating them and subsequently transforming them in canonical model based on FHIR (Figure 1). The aim is to use such FHIR complaint information components in provision of knowledge services for clinicians, such as deriving clinical quality measures and, in some cases, creating new insights (e.g. ML, AI and Clinical Decision Support systems), which are then recorded as new FHIR resources. Such normalised data or newly created data can then be linked as part of clinical workflows and made available via portals or dashboards to clinicians and patients (engagement layer), while respecting security, consent, privacy, ethics, and other policies (specifying as part of Data Access services layer).

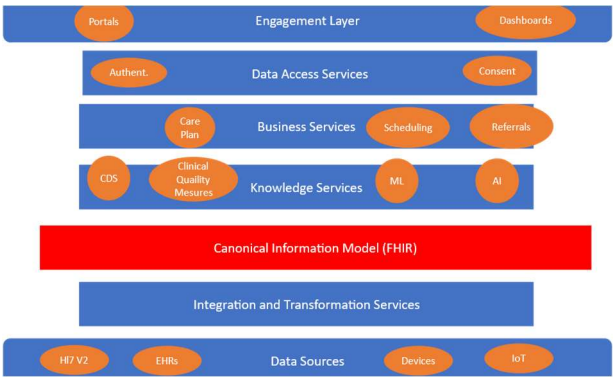

Figure 1: A Digital Health Ecosystem

There are many integration points in such a digital health ecosystem where analytics or AI services can be inccluded. For example, real-time analytics systems can be integrated to intercept source data from laboratory systems and apply business rules, such as duplicate orders, data quality assurance or clinical decision as was reported in [8]. In addition, these rules can be augmented with ML rules to support continuous adjustment of training sets in real time. Similarly, one can access data from IoT devices and apply specific rules either on-fly or once the data are stored in appropriate repositors for subsequent processing. These however require ability to develop interceptors for messaging systems.

The use of FHIR canonical model on one hand adds to the common understating of information models in an ecosystem and, on the other hand, serves as the foundations for many specialised vendors to provide value-add knowledge services, such as Clinical Decision Support (CDS), analytics or ML/AI, while accessing data repositories such as electronic health records or even legacy systems stores. The common understanding associated with the standard representation of healthcare concepts in FHIR resources, is also key for supporting patients in defining their privacy and consent preferences for providers in accessing their data, at any level of granularity as required. This can be supported using the FHIR consent resource.

It is to be noted that FHIR can be regarded as a logical model, with the resources used for specific applications deployed within cloud or distributed system environments.

III. CONSENT– FOR USER CONTROL AND TRUST

Consent mechanism is increasingly becoming an important element in supporting individuals in controlling access to their personal information, such as per GDPR regulations in Europe [14]. In healthcare, this is also a result of trends towards increasing patient awareness and empowering the patient in the healthcare process, as for example dictated by the US 21st Century Cure act [13].

Consent is an important element in the context of AI applications as well. For example, individuals may want to define how access to their healthcare information is used by AI systems. This is to ensure that this information is used as a means of helping to improve their own healthcare, or healthcare of others (e.g. through involvement in research studies), but they may wish to specify controls to protect them against misuse of this information for other purposes, such as for commercial interest of AI solution vendors.

The above examples suggest a value in the transition of the usual, paper expression of consent into *computable* representations. One such initiative is a recent HL7 standardisation effort on Consent Management Service [11], which is a platform independent model (PIM) for specifying consent. This PIM can be applied to any specific platform, such as for example FHIR consent resource [5]. This PIM model was influenced by several commercial and standardisation efforts, the latter of which is the RM-ODP standard [1]. It was also influenced by our recent work on modelling computable expression of enterprise policies, based on the use of deontic and accountability concepts, as presented in [12].

Section III.A provides a summary of our approach for computable expression of policies, as introduced in our earlier work [12], and its application to modelling consent. This is essentially part of organisational interoperability concerns introduced earlier, and it is to be noted that the proposed policy model is generic and can be used for the computable expression of ethics (see section IV).

A. Generic computable model for enterprise policies

The specification of enterprise policies can be expressed in terms of constraints for the actions of the parties who participate in interactions. These constraints are typically prescribed by an external authority, e.g. a legislative, jurisdictional or regulative context, such as, HIPPA [39], US Final Cure Act [13] and GDPR [14]. A policy context thus consists of rules prescribed by the applicable

jurisdictions or organizational entities, and we use the RM-ODP concept of community, domain and federation [1] to model this context, as elaborated next.

1) Policy context – community model

A community defines how a set of participants should behave in order to achieve an objective. These participants (or enterprise objects in RM-ODP terms) fulfill roles in a community, and thus accept policy constraints that apply to the roles, as stated in the contract for community [20]. At any point in time, at most one enterprise object can fulfil a community role, but an enterprise specification may include a number of roles of the same type, each fulfilled by distinct enterprise objects, possibly with the constraint on the number of roles of that type that can occur, for example, maximum number of patients in a ward. Note that most enterprise objects display behaviour and are thus referred to as active enterprise objects, the special kind of which is party, with legal responsibility and accountability, as introduced in section III.A.2).

A community role can thus be played by a party, which models a natural person or legal entity, and its behaviour is constrained by the behaviour specified by that role. A role in a community can also be played by another community, making it possible to model hierarchical policy contexts.

A specific type of community suitable for the defining policy contexts is domain community (or simply a *domain*). A domain can be used to model legal or regulative contexts for which a particular controlling object, for example, legal or regulatory authority, prescribes a set of policies that define legal or regulative constraints for individual members of that domain. Examples are obligations, prohibitions or permissions defined by the General Data Protection Regulation (GDPR) authorities and the controlling objects are the so-called Data Controllers [14]. Another example of the controlling object is that of National Data Protection Authority, tasked with protecting information privacy [42].

Domains can be arranged hierarchically, through subdomains, which are subsets of a given domain, but can also be federated. Expressing *federation* is important for healthcare interoperability in view of the need to manage the combined actions of private and public stakeholders within health sector and across other sectors [19].

2) Policy constraints - deontic concepts

There are three fundamental types of policy constraints that reflect rules of any normative system, namely obligations, prohibitions and permissions [4]. Their formal expression is the subject of deontic logic [18] and these are often referred to as deontic constraints.

An *obligation* is a prescription that a particular behaviour is required. An obligation is fulfilled by the occurrence of the prescribed behaviour.

A *permission* is a prescription that a particular behaviour is allowed to occur. A permission is equivalent to there being no obligation for the behaviour not to occur.

A *prohibition* is a prescription that a particular behaviour must not occur. A prohibition is equivalent to there being an obligation for the behaviour not to occur.

The above definitions have been the subject of standard deontic logic [18], but their application in enterprise distributed computing requires explicit association with the agent to which these constraints apply. This is also needed to consider an agent's goal-seeking behaviour, which may result in their willingness to violate the policies with the expected benefit of reward from doing so.

The way that deontic constraints are associated with the agents is through the concept of *deontic tokens [4]*. These are enterprise objects which encapsulate deontic constraint assertions. The holding of the deontic tokens by active enterprise objects constrains their behaviour. This modelling approach provides a basis for manipulating deontic tokens, for example, passing them between parties to model delegations, and activation or de-activation of policies that apply to the active enterprise objects. There are three types of deontic tokens that represent deontic constraints. These are called burden, representing an obligation, permit, representing permission and embargo, representing prohibition.

In the case of a *burden*, an active enterprise object holding the burden must attempt to discharge it either directly by performing the specified behaviour or indirectly by engaging some other object to take possession of the burden and perform the specified behaviour. In the case of *permit*, an active enterprise object holding the permit is able to perform some specified piece of behaviour, while in the case of *embargo*, the object holding the embargo is inhibited from performing the behaviour [4].

In order to support the changes in policies that apply to active enterprise objects, the concept of a *speech act* is introduced. This is a special kind of action that is used to modify the set of tokens held by an active enterprise object. The name was chosen by analogy to the linguistic concept of speech act, which refers to something expressed by an individual that not only presents information but performs an action as well [27]. So, a speech act changes the state of the world in terms of the association of deontic tokens with active enterprise objects, such as patient giving permit to a researcher to access their health record.

Deontic constraints and tokens provide foundations for expressing many types of policy constraints across enterprise objects in a system, including both human actors and automated agents, such as AI systems. ODP-EL provides added formalism to express traceability of obligations of parties, according to their broader responsibilities derived from ethical, social or legal norms [20]. This formalism is referred to as *accountability* concepts, as described next.

3) Policy constraints - accountability concepts

We have informally introduced the concept of *party* in III.A.1). Formally, party is defined as an enterprise object which models a natural person or any other entity considered to have some of the rights, powers and duties of natural person, for example, company [4]. ODP-EL introduces two other concepts which are useful to describe many forms of delegation in enterprise systems. *Principal* is defined as a party that has delegated something (e.g. authorization or provision of service) to another, and *Agent* is defined as an active enterprise object that has been delegated something (e.g. authorization, responsibility of provision of service) by, and acts for, a party (e.g. in exercising the authorization, carrying out responsibility).

Delegation is an action that assigns something (e.g. authorization, responsibility of provision of service) to another object. It is through this mechanism that deontic

tokens can be passed across different active enterprise objects, with one example being a delegation from principal to agent, as mentioned above. Delegation is one action type in ODP-EL related to accountability, but there are several other action types to capture important business events in any organizational system, and reflect the dynamics of communication amongst parties, and broadly, active enterprise objects. These action types are listed next [4].

Commitment is an action resulting in an obligation by one or more participants in the act to comply with a rule or perform a contract. This effectively means that they will be assigned a burden. Examples include commitments by clinicians to deliver safe, reliable and effective healthcare.

Declaration is an action by which an object makes facts known in its environment and establishes a new state of affairs in its environment. This can, be performed by an AI system (or a party managing it), for example, informing the interested parties about the result of some analysis.

Evaluation is an action that assesses the value of something, which can be in terms of variables such as importance, preference, and usefulness. In digital health, variables can be performance parameters used, through research applications for example, to either express administrative performance or some accuracy or reliability measures. They can be used to assess the fairness of training data or as part of mechanisms to measure the impact of AI algorithms as part of their explainability requirements [12].

Prescription is an action that establishes a rule. Prescriptions provide a flexible and powerful mechanism for changing the system's business rules at runtime, enabling dynamic adaptation to respond to business changes and new needs. This ability is important in any digital health system, to establish the applicability of new policies, such as reflecting new legislations, or after the adoption of recommendations from AI system components [12].

Authorization is an action indicating that a particular behaviour shall not be prevented. Unlike a permission, an authorization is an empowerment. In terms of deontic tokens, the enterprise object that has performed authorization will issue a required permit and will itself undertake a burden describing its obligation to facilitate the behaviour. For example, the authorization for the consumer to challenge AI decisions is giving them permit to do so by the AI system (or its creator/manager) who has the burden to do so [16].

B. Modelling privacy consent

This section illustrates how the generic policy concepts introduced above can be used to represent policy constraints associated with privacy consent.

The privacy consent is modelled using the key roles of Grantor and Grantee, supported by several other roles needed for consent management.

These roles are thus part of the privacy consent community, which are described next, together with several applicable deontic and accountability constraints, and as initially proposed in [12].

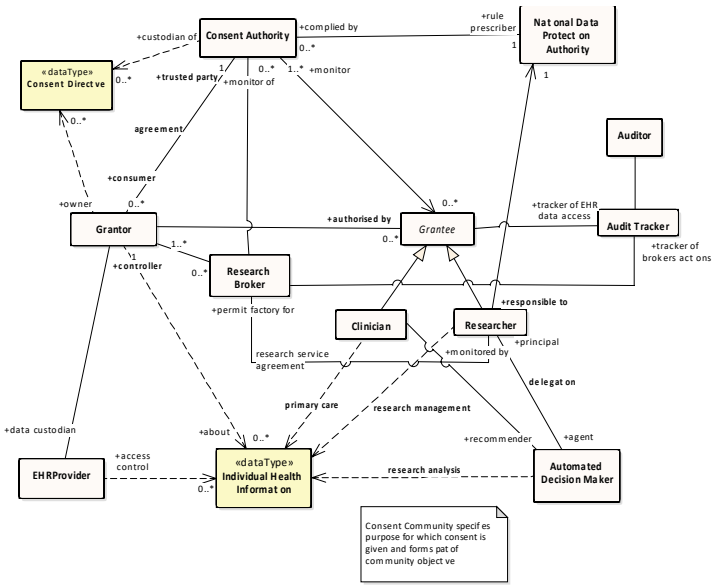

Figure 2: Privacy consent management community

1) Privacy consent community – key roles

This community specifies the following role types:

- Grantor (Figure 2), to be fulfilled by any individual giving consent under a set of permission rules, reflecting some competence criteria, such as being of legal age, having normal cognitive function etc.
- Grantee, to be fulfilled by professionals with the required credentials, namely:
  - Clinician, with permission to access Grantors' individual health information for care purposes, covered by the patients consent for primary care, e.g. access to all of the patient information in an emergency situation, with certain constraints, such as time period from the emergency event.
  - Researcher, with permission to access Grantors de-identified health data for research purposes and obligation not to perform re-identification of patient data, as prescribed by National Data Protection Authority.
- *Consent Authority*, a trusted party responsible for storing individuals' consents and overseeing the consent agreement rules; it can also facilitate ethics approvals for the secondary use of data.
- Research Broker, a legal entity authorized to search patient health data and consent data to identify patients suitable for research projects. The Broker is responsible to ensure that patient preferences are enforced. It is accountable to the Consent Authority and the National Data Protection Authority.
- National Data Protection Authority, responsible for defining and enforcing data protection policies, as legislated [42].
- *Electronic Health Record (EHR) provider*, who is custodian of individuals' personal health data in their EHR records.
- *Automated Decision-Maker*, performing analytics, recommendations and in some cases, active decision-making; this role guides and augments activities of clinicians, researchers, and other stakeholders, such as

population health experts; this role can be fulfilled by clinical decision support systems or AI systems.

- *Audit Tracker*, which logs events associated with actions of clinicians and researchers to generate audit trails, which can be used for subsequent activity analysis, such as performed by an Auditor, listed next;
- *Auditor*, who provides analysis of event traces produced as above to support performance analysis or forensic investigations, to detect breaches and their consequences; an example is detecting breaches of clinicians accessing healthcare records outside of them providing care, or researchers accessing linked data provided by third parties, both of which are forbidden.

2) Deontic constraints

The following are examples of deontic constraints that apply to the roles of the privacy consent community [16]:

• Permission of the Grantor given to the Consent Authority to store consent agreements, for example, valid for a specified time period defined by the Grantor.

• Permission of the Grantor to the Broker to search patients' data and if it satisfies researcher criteria include a link to this data in a data set for the researcher.

• Obligation on the Audit Tracker to log data access by the Grantee reliably and on-time and provide access to the audit trail by the Auditor; the tracker may also have an obligation to log actions of Research Broker which may be needed for forensic purpose.

• Authorization of the Grantor to the Grantee to access the Grantor's individual health information; this is realized through this chain of authorization:

– Grantor issues permit to the Research Broker for searching their data to establish whether they satisfy research question criteria.

– Research Broker issues a research permit to the researcher which includes a list of Grantors that provided consent to access their de-identified health data and whose data satisfy the research question.

– EHR provider provides access permit to the researcher to access health records of specific patients, provided researcher has credentials requested by the EHR provider.

3) Accountability constraints

Authorization is modelled using a combination of permit and burden deontic tokens. For example, authorization of the Grantor to the Broker involves both the permit being passed from the Grantor to the Broker to search its record but also places an obligation on the Grantor itself, through the corresponding burden, to ensure that access to its record is ultimately enabled. This authorization action is also a speech act because it changes the deontic state of both the Grantor and Grantee. The effect of this speech act is that the existing Grantor's permit to the Broker to search its healthcare data is passed on to the Grantee. In this example, we assume that the consent directive gives permission to the Researcher to access the Grantors health data but prohibits access to the Grantor's mental health data (if it exists).

The use of speech acts and deontic tokens provides flexibility in describing the dynamics of deontic constraints and passing of tokens, including to the parties with ultimate legal responsibility. For example, many data protection rules defined by a National Data Protection Authority set accountability and legal responsibility expectations for actions of researchers involved in using grantor's data. These data protection rules were established through *prescription* actions), performed by the National Data Protection Authority, which essentially establishes obligations and permissions for all the parties involved in accessing patient data.

IV. ETHICS – FOR RESPONSIBLE DATA ACCESS AND SHARING

The expression of privacy consent, as a vehicle for patients to state their desires for controlling the use of their personal health information, as often considered as part of a broader ethics framework, which is about the doings of 'rights' or 'wrongs' in the context of delivering healthcare.

Ethics challenges have become more prominent in recent times due to potential concerns associated with the use of AI. These include the impact of AI decisions on patient care, without full ability to interpret AI algorithm decision making process, or how to attribute accountability to such decisions in case of clinicians' use of it. There are increasing efforts to develop *ethics principles* to guide the design and implementation of AI enabled systems in general [15], and with specific digital health focus [17].

We believe that a computable expression of *ethics principles* and *ethics concepts* can be a useful tool for practitioners faced for designing and deploying AI systems as part of their digital health enterprise. The following provides a summary of our approach and recommendations as initially proposed in [16].

Firstly, we proposed a structured approach for progressive refinement of *ethics principles* into formal models that can support reasoning about, designing, implementing and running AI-enabled digital health systems. We refer to this as 'ethics-by-design' methodology and again, we use a deontic-based formalism as a common theme for two aspects of the refinement problem. At the *analysis* level, we use it to facilitate the expression of the ethics principles to guide what the *system* should or should not do. At the *design* level, we use it for precise expression of behaviour constraints of *actors* involved in a digital health system, including their accountability, as introduced in the previous section.

The following set of ethics principles are identified for consideration when embarking on any digital health projects, as summarised in [16], and based on [10] and [17]:

- Privacy data protection – must ensure that people's private data are protected and prevent breaches that could cause any damage to people
- Accountability – should identify people and organisations responsible for the design and implementation of digital health systems, including AI
- Compliance – must comply with relevant international, national, regulatory and legislative frameworks
- Safety and reliability – must ensure that systems are designed to avoid any negative impact to consumer
- Fairness – must ensure the training data for machine learning is free from bias that might cause the algorithm to behave unfairly against individual or groups.
- Explainability – must inform consumers about how exactly their data is used by an AI system and how it makes decisions
- Contestability – must allow consumers to challenge the output of the AI algorithm when it impacts them

- Do no harm – must not be designed to harm or deceive people through its decisions.

1) Ethics-based analysis

The ethics principles express rules that specify the expected properties of a digital health system to reflect ethics requirements. These rules can be treated as the deontic modalities that apply to the *system*, considered as an entity performing actions, including decision making that can affect consumers. These in turn can serve as an input to the detailed design and run-time enforcement which includes deontic constraints that apply to the *participants* involved in the system.

Consider first the *privacy protection* principle. This can be modelled as an obligation of the system to respect privacy constraints for accessing personal health data, as specified by the consumers' consent. This obligation can then be refined into a number of fine-grained, design-level constraints, on actions of agents involved in controlling and accessing private health information. This includes individuals who specify consent rules (grantors), clinicians/researchers who access personal information for the primary/secondary use purpose (grantees), and authorities involved in the governance over the use of personal data.

The *compliance* principle can be interpreted as an obligation of a system to respect the applicable regulation and legislative rules, such as the Privacy Act in Australia and related regulations to do with secondary use data.

The *accountability* principle can be regarded as an obligation for a digital health system to identify *parties* legally responsible for the creation and deployment of the system. The ability to clearly represent chains of accountability and responsibility, including the links to appropriate legislative and regulatory authorities, increases consumers trust, in particular in AI enabled systems.

Safety and reliability, has been a core principle of the development of medical technologies for quite some time [17], referring to the obligations of medical devices and clinical systems (i.e. their providers) to deliver services in a way that is unlikely to cause danger, risk, or injury to individuals. This is directly related to the *DoNoHarm* principle identified in [10] as an AI ethics principle, which states that "civilian AI systems must not be designed to harm or deceive people and should be implemented in ways that minimise any negative outcomes". This can be modelled as a prohibition of the system to create an algorithm that could cause harm to the application consumer. This prohibition should be then traced back to the obligations of the system creator not to design such a system, which is again manifested in their accountability, typically delegated to their organisation. Note that machine learning algorithms do not provide safety and reliability guarantees typical in safety critical systems such as pacemaker devices. This is because of their inherent stochastic nature, and further research is required to better position AI solutions in the context of such guarantees. A recent direction is in combining machine learning with automated reasoning techniques to support building explainable and dependable AI systems [34].

The *explainability* principle of an AI system is an obligation of an AI system to provide information to the users about how the AI algorithm makes a decision and which data set it is using to do so. This is of particular importance when such systems are used for clinical decision making to augment the work of clinicians. There are several techniques that assist in explaining AI's decision-making process, of which the LIME method (Local Interpretable Model-Agnostic Explanations) attracted a lot of attention recently[31]. Further, some authors propose the use of blockchain to track all the stages in AI algorithms as a way of understanding decision making processes. Such blockchain-based trails can assist to determine whether humans (and who specifically) or machines are at fault in case of accidents [28].

The *contestability* principle can be expressed as an obligation of an AI system to allow (i.e. give permission to) consumers to challenge the use or output of the algorithm. This permission can also be considered as an authorisation given to the consumer to participate in the challenge process.

2) Ethics-based design

The system level deontic statements presented in the previous section can be translated into detailed behavioural constraints for the actions of the parties and system components involved, according to the ODP-EL standard, introduced.

Figure 3 depicts the applicability of deontic and accountability constraints on automated and legal entities and their links with the development methodology, of which analysis and design were discussed above. A specific development environment would dictate a set of tools for build and run phases. UML-based model-driven tools with UML profile for ODP support can be a potential candidate [20], integrated with specific AI platforms. Further, specific technology platforms can be used, and they can impact the selection of controls, such as the some distributed ledgers, which provides new solutions for the implementation and protection of digital identifiers [26].

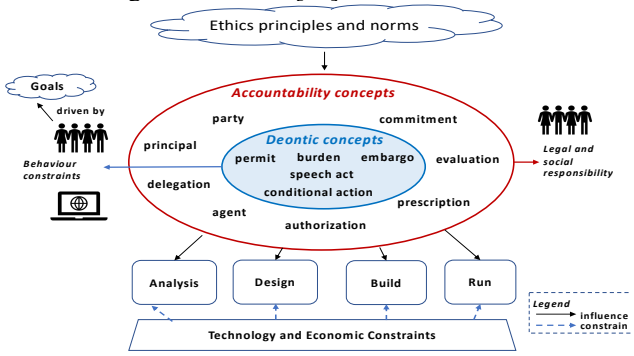

Figure 3: Ethics aware development methodology

V. CASE STUDY – PRACTICE MANAGEMENT SYSTEMS

This section provides an example of how the concepts introduced in previous sections can be applied in the context of a digital health ecosystem focused on primary care provision in Australia. We are leveraging current architecture approaches with the aim to position future requirements to better support patient-facing and patient-controlled support. The focus is on supporting the integration of consent and ethics computational components, while leveraging the emerging interoperability approaches and technologies based on FHIR.

A. Consent

The *organisational* concerns are captured by the objectives, key roles and policies of the privacy consent management community introduced in Figure 2. In the context of primary care, a patient typically has their own general practitioner (GP), using a Practice Management Software (PMS) system to support capturing clinical and administrative information about patients in their own electronic health records (EHR)[1]. Further, in relation to the roles of the consent management community, a patient fulfills the role of a Grantor, GP fulfills the role of Grantee, and a PMS provider fulfills the role of EHRProvider. Patient Individual Health Information (IHI) is typically stored in a database record of a PMS, which can be either a desktop system or increasingly deployed on a cloud-based platform. Note that in Australia, it is the Office of the Australian Information Commission, acting as national data protection authority for Australia [42].

The *information* concerns are defined by an information model, the elements of which capture key clinical information components, such as observation, medication, but also administrative concepts such as patient demographics, appointments, payments. These concepts are typically implemented using columns in a relational database in PMS systems but are increasingly migrating into component and service-based model, in particularly when deployed in the cloud environment. One such approach is the use of HL7 FHIR standard to represent information models, using FHIR resources as introduced in section II.C.

FHIR is also providing a computational expression for consent specification through the FHIR Consent resource. This resource is "a record of a healthcare consumer's choices, which permits or denies identified recipient(s) or recipient role(s) to perform one or more actions within a given policy context, for specific purposes and periods of time." Much of the data elements in FHIR consent resource can capture the platform independent policy specification presented in section III.B, but the details of this mapping go beyond this paper. It is to be noted that there is a strong alignment between FHIR Consent resource and the platform independent consent specification recently published by HL7 International [11], which in part was influenced by the consent policy model specified in section III.B of this paper.

The *technical* interoperability concerns are supported by a flexible open API based architecture. For example, the use of FHIR compliant technical infrastructure, most notably FHIR servers, with FHIR defined API interfaces, supports standard way of exchanging data or composing applications, supporting fast deployment and management of applications in an open digital health ecosystem. A FHIR server can be provided by a vendor that supports GP practices, implementing an electronic health record repository. This server can be deployed in a cloud environment and can also include a separate repository of Consent Directives. The security requirements can determine whether such repository can be within the tenant of the EHR provider or under control of a separate trusted party, such as Consent Authority, introduced in Figure 2. The use of cloud based FHIR solutions allows integration of traditional PMS solutions with broader digital health ecosystem, including for example with hospitals, in support of referrals and discharge, residential aged care facilities, allied health and community providers, pharmacy organisations and so on. This also includes support for building patient-facing portals, through which patients can better engage with their GPs, including defining and updating their consent rules.

It is key that FHIR standard provides basis for many integration points, but these need to be governed by business, information and technical agreements, including the security, consent and ethics policies.

B. Ethics

In the context of primary care, the following scenarios will require a systematic approach to ethics aware design and implementation:

- use of patient personal health information for the purpose of secondary research
- use of analytics and AI based solution to help GPs in making better evidence-based decision, including the use of Clinical Decision Support
- third parties' use of patent information, either provider or patient entered, through their portals

In the context of PMS systems, Research Brokers facilitate matching of research interests of research organisations (ether for clinical research or for statistics-focused research done by government organisations), with the de-identified patient data available from GP practices. In this case an EHRProvider is obliged to ensure that the data shared with the Brokers are de-identified, although this obligation might be delegated to the Broker, if governed by a separate partnering agreement. This implementation is required to satisfy the *privacy protection* ethics principles, although finer grained controls can be specified by the patient themselves.

The use of Automated Decision Makers by a GP is typically in terms of rule-based systems, as in many CDSs, some of which have real-time support, although there is an increasing interest in the use of analytics and AI systems. In this role for example, analytics applications can be used to better understand different clinical patterns associated with different patient cohorts, and this functionality can be part of an internal PMS component, although a separate third-party can be contracted, making sure that all relevant consent and ethics principles mentioned in section IV.1) are satisfied, and by applying the ethics policies, as illustrated in IV.2). Another example is the use of AI solutions to guide GPs in making faster clinical decision processes, but for a scalable use of AI in this context, some of the legal issues associated with accountability ethics principle need to be addressed as also highlighted in [17].

VI. DISCUSSION

The focus of this paper is on the solution approaches for scalable AI, with particular focus on the interoperability, consent and ethics, of particular relevance for digital health. There are other factors that also need to be considered which are briefly discussed next.

[1] PMS systems are offered by several vendors in Australia, one of which is Best Practice Software [40].

One concern is how to deliver quality AI solutions while taking into account the computational cost. There need to be a balance between obtained AI model accuracy and economic, environmental, and social cost of reaching the accuracy level, such as for example when preforming intensive natural language processing (NLP) applications [38].

Another concern is looking at scalable management of data and models. The former is about the careful collection and curation of data sets for use in AI systems, to address challenges associated with time consuming, expensive, error-prone, and labour-intensive [6] aspects of data collection. The latter one is about reusing the models developed through training over a specific dataset, but for different problem of target domain, one technique of which is referred to as transfer learning.

Further concern is how to understand how an AI system operates, or explainable artificial intelligence (XAI). XAI refers to a set of techniques available to human users to provide them with transparency and trust about the results and output created by machine learning, such as with deep learning and neural networks. Explainability can help developers ensure that the system is working as expected, it might be necessary to meet regulatory standards, or it might be important in allowing those affected by a decision to challenge or change that outcome. Explainable AI is crucial for an organization in building trust and confidence when putting AI models into production [9]. In general, investigating model behaviours through tracking model insights on deployment status, fairness, quality and drift is essential to scaling AI [9].

Depending on the complexity of the model, one can use 'white box' models, that are transparent and easily interpretable, providing relatively straightforward explanation how models provide predictions and what are the influencing variables, i.e. how the models behave. For example, simple decision trees or ordinary regression with a few variables, make it easy to tell how the variables combine to form the system's predictions.

On the other hand, 'black box' models refer to situations where simple models are not sufficient to explain a particular machine learning activity, making it much more difficult, or even impossible, for ordinary humans to understand how an algorithm makes a decision. Examples are neural networks with many layers or convolutional neural networks. In these cases, the explanations of how black box models behave are supported by applying a second (white box) algorithm, developed to approximate the outputs of the black box. This second algorithm is trained by fitting the predictions of the black box and not the original data and is used to develop post-hoc explanations for the black-box outputs and not to make actual predictions.

In spite of their increasing popularity as a research topic, there are some challenges in practical use of black box based XAI for digital health, as recently discussed in [10], and practitioners would need to be aware of these. The authors first consider an example of a point scoring system used by many doctors for calculating patients' heart disease or stroke risk based on their blood pressure, cholesterol levels, age, and other characteristics. This is an interpretable AI, providing transparency and helping one understand how a model works. It is simple, intuitive, and easy to grasp. Their argument however is that, in many medical applications where developers need a more complicated 'black box type' model, there may be certain issues with the correctness of explanation. So, in the stroke risk example, the white-box explanatory algorithm might tell a patient that their high risk of stroke, as it was predicted by the black-box model, is consistent with a linear model that relies on their age, blood pressure, and smoking behaviour. But it is easy to imagine many other explanations that can be generated that are also consistent with the black-box prediction. For example, the patient's risk of stroke could also be consistent with a decision tree that relies on one's gender and diabetes status instead of blood pressure and smoking status [10].

These issues associated with imperfection of XAI may suggest that in some cases, regulators should consider alternative methods for AI product assurance, such as the use of clinical trials. This was indeed what the authors propose.

VII. Conclusions and Future Work

One goal of the paper is to provide guidance to practitioners interested in the use of ML/AI in addressing specific digital health challenges. Another goal is to offer new insights to the ML/AI researchers about broader aspects of digital health ecosystem and help them in positioning of their AI/ML solutions in such systems. The aim here is to stimulate many types of collaboration between academic and practitioners, towards developing a learning health system.

Regarding the first goal, the paper has first identified relevant interoperability frameworks and standards and then proposed a computable policy framework to support integrating consent and ethics rules in support of analytics and AI in digital health ecosystem. These can be used by practitioners as guidance when embarking on building AI applications, for discrete use cases or on end-to-end basis.

We plan to test this framework in several concrete analytics or AI projects, and where necessary update these for future use. For example, it would be of interest to consider how fine grain consent rules can be integrated as part of a specific patient management system's portal, allowing patients to define control over their personal health information, while the portal still being managed by the provider. This would complement current consent solutions which are focused on specifying consent for receiving communication, such as permitting receipt of pathology reports but excluding receipt of advertising emails. A related issue is the level of adoption of such expressive framework by consumers, which would reflect various demographics aspects of consumer, which can be informed by recent analysis presented in [41].

Regarding the second goal, it would be of value to consider how one can get AI to scale within a particular healthcare organisation or collaborative structure. One can start with reimagining their own business processes, or function enabled by AI end to end, and incrementally adopt the use of AI, leveraging lessens from previous efforts, while making use of the toolbox of enabling technologies identified in this paper.

We also plan to investigate the role of distributed ledgers and digital twins, which can provide their own components as part of the digital health ecosystem, with links to analytics, ML and AI applications.

In terms of potential future work related to ethics, we are interested in investigating tool-based support for relating ethics principles into design of digital health systems. This could involve the use of existing UML tooling, including support for UML for ODP standard [25], but also experimenting with broader set of tools and relevant formalisms for research purposes, including the use of ontologies to represent the ethics concepts discussed in section IV.2). One specific topic is investigating how the concept of value can be modelled in support for reasoning about ethical dilemmas and conflicts. For example, we are planning to look at the *possible word semantics*, based on *Kripke model*, augmented with the concept of *utility*, as also mentioned in the ODP-EL standard [4] .

ACKNOWLDEGMENT

I would like to thank anonymous reviewers for their comments to an earlier version of this paper. I would also like to thank my colleagues from Best Practice Software, especially Dr Frank Pyefinch, Dr Fabrina Hossain, William Dunford and Peter Polacek, for providing valuable input to the paper, regarding the deployment concerns within the Practice Management Software (PMS) application domain.

REFERENCES

[1] Azure Machine Learning, https://azure.microsoft.com/en-au/services/machine-learning/

[2] Machine Learning on AWS, https://aws.amazon.com/machine-learning/

[3] Google Cloud AI, https://research.google/teams/cloud-ai/

[4] ISO/IEC 15414, *Information technology: Open distributed processing, Reference model – Enterprise Language*, 3rd ed, 2015.

[5] Hl7 FHIR, https://www.hl7.org/fhir

[6] Scalable AI, Carnegie Mellon University, SEI, https://resources.sei.cmu.edu/asset_files/WhitePaper/2021_019_001_735330.pdf

[7] HealthConourse, http://phaseone.net/sites/default/files/2019-07/HealthConcourse_Overview_113018.pdf

[8] Berry, A., Milosevic, Z. 2013. Real-time analytics for legacy data streams in health: monitoring health data quality. In EDOC 2013.

[9] Explainable AI, https://www.ibm.com/au-en/watson/explainable-ai

[10] Boris Babic and Sara Gerke, Explaining medical AI is easier said than done, https://www.statnews.com/2021/07/21/explainable-medical-ai-easier-said-than-done/, Stat, June 21

[11] HL7 Services Functional Model: Consent Management Service, Release 1, July 2021.

[12] Z. Milosevic, Enacting policies in digital health: A case for smart legal contracts and distributed ledgers?, The Knowledge Engineering Review, 35, Cambridge University Press, Feb. 2020

[13] 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program, https://www.healthit.gov/curesrule/

[14] GDPR, General Data Protection Regulation, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX: 32016R0679ss

[15] Dawson D, Schleiger E, Horton J, McLaughlin J, Robinson C, Quezada G, Scowcroft J, Hajkowicz S (2019) *Artificial Intelligence: Australia's Ethics Framework*. Data61 CSIRO.

[16] Z. Milosevic, Ethics in Digital Health: a deontic accountability framework, Proceeding of EDOC2019 conference

[17] Microsoft, *Healthcare, artificial intelligence, data and ethics* - A 2030 vision, Dec 2018.

[18] G.H. von Wright, *Deontic Logic*, Mind, Vol 60, pp. 1-15, 1951

[19] Z. Milosevic, A.Bond, *Digital health Interoperability frameworks: use of RM-ODP standards*, IEEE EDOC SoE4EE workshop, 2016.

[20] P.F. Linington, Z. Milosevic, A. Tanaka and A. Vallecillo, *Building Enterprise Systems with ODP, An Introduction to Open Distributed Processing*, Chapman & Hall/CRC Press, 2011.

[21] Australian Institute of Health and Welfare 2018. *Australia's health 2018. Australia's health series no. 16, Secondary Use of Health Information, section 2.5*, 2018.

[22] S. K Ludwin, T . Murray, *Dilemmas in medical ethics in the age of big data*, Multiple Sclerosis Journal, 2017, Vol. 23(10) 1306–1308

[23] https://plato.stanford.edu/entries/ethics-deontological/

[24] www.zdnet.com/article/is-it-moral-to-benefit-from-research-while-opting-out-of-electronic-health-records/

[25] P. Linington, H. Miyazaki, A. Vallecillo, *Obligations and Delegation in the ODP Enterprise Language*, the IEEE 16th International Enterprise Distributed Computing Workshops, 2012.

[26] G. Zyskind, O. Nathan, A. Pentland, Enigma: *Decentralized computation platform with guaranteed privacy*, 2015.

[27] https://en.wikipedia.org/wiki/Speech_act

[28] T. N. Dinh, My T. Thai, *AI and Blockchain: A Disruptive Integration*, IEEE Computer, vol. 51 no. 9, Sept 2018.

[29] J. Van den Hoven, G.J Lokhorst, *Deontic Logic and Computer - Supported Computer Ethics*, Metaphilosophy, Jan 2003

[30] Van den Hoven, J., Miller, S., & Pogge, T. (Eds.). (2017). *Designing in Ethics*. Cambridge: Cambridge University Pres

[31] M. T. Ribeiro, S. Singh, C. Guestrin *"Why Should I Trust You?" Explaining the Predictions of Any Classifier*, Proc. the 22nd ACM SIGKD Int. Conf. on Knowledge Discovery and Data Mining, 2016

[32] Dastani, M., Torroni, P., & Yorke-Smith, N. (2018). *Monitoring norms: A multi-disciplinary perspective.* The Knowledge Engineering Review, 33, E25. doi:10.1017/S0269888918000267

[33] Laponte, C, Fishbane, R., *The Blockchain Ethical Design Framework*, Georgetown University

[34] H. Bride, J. Dong, JS Dong, Z. Hou, *Towards Dependable and Explainable Machine Learning Using Automated Reasoning*. ICFEM 2018, pp 412-416.

[35] https://www.computerworld.com.au/article/661996/australian-businesses-split-over-where-ai-accountability-lies/

[36] C Griffo, JPA Almeida, G Guizzardi, Conceptual Modeling of Legal Relations, Int. Conf. on Conceptual Modeling, 2018, pp. 169-183.

[37] Forrester Total Economic Impact™ (TEI) of Azure Machine Learning, https://azure.microsoft.com/en-ca/resources/forrester-total-economic-impact-tei-of-azure-machine-learning/

[38] R. Schwartz, J. Dodge, N. A. Smith, O. Etzioni, Green AI, Communications of the ACM, Dec 2020, Vol. 63 No. 12, Pages 54-63

[39] https://www.hipaajournal.com/hipaa-compliance-checklist/

[40] Best Practice software, https://bpsoftware.net/

[41] Sustaning The Growth Of Digital Health, Accenture Health Consumer Survey, Australia Findings, 2020

[42] National data protection authority, https://en.wikipedia.org/wiki/National_data_protection_authority