PHAX: A Structured Argumentation Framework for User-Centered Explainable AI in Public Health and Biomedical Sciences

Bahar İlgen[1,*], Akshat Dubey[1,2] and Georges Hattab[1,2]

[1]Center for Artificial Intelligence in Public Health Research (ZKI-PH), Robert Koch Institute, Nordufer 20, Berlin, 13353, Germany
[2]Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 14, Berlin, 14195, Germany

Abstract

Ensuring transparency and trust in AI-driven public health and biomedical sciences systems requires more than accurate predictions—it demands explanations that are clear, contextual, and socially accountable. While explainable AI (XAI) has advanced in areas like feature attribution and model interpretability, most methods still lack the structure and adaptability needed for diverse health stakeholders, including clinicians, policymakers, and the general public. We introduce PHAX—a Public Health Argumentation and eXplainability framework—that leverages structured argumentation to generate human-centered explanations for AI outputs. PHAX is a multi-layer architecture combining defeasible reasoning, adaptive natural language techniques, and user modeling to produce context-aware, audience-specific justifications. More specifically, we show how argumentation enhances explainability by supporting AI-driven decision-making, justifying recommendations, and enabling interactive dialogues across user types. We demonstrate the applicability of PHAX through use cases such as medical term simplification, patient-clinician communication, and policy justification. In particular, we show how simplification decisions can be modeled as argument chains and personalized based on user expertise—enhancing both interpretability and trust. By aligning formal reasoning methods with communicative demands, PHAX contributes to a broader vision of transparent, human-centered AI in public health.

Keywords

Explainable AI, Argumentation-based Explainability, Structured Argumentation, User-Adaptive Explanation, Public Health Informatics, Natural Language Processing, Trustworthy AI, Health Communication

1. Introduction

As artificial intelligence (AI) becomes increasingly embedded in public health systems, ensuring that AI outputs are understandable, trustworthy, and tailored to diverse stakeholders has become a critical challenge [1, 2, 3, 4]. Moreover, recent calls in public health literature highlight the necessity of Explainable AI (XAI) to foster transparency and professional trust in healthcare applications. From clinical diagnostics to vaccination policy, AI now plays a role in high-stakes decisions that affect patients, practitioners, and entire populations. Applications in areas like pandemic preparedness have made clear that epidemiological decision-making increasingly depends on the integration of XAI [5]. Yet, the logic underlying many AI-driven decisions often remains obscure, fueling concerns over accountability, fairness, and interpretability.

The goal of XAI is to address such issues by shedding light on model behavior. However, most existing XAI approaches—such as feature attribution or counterfactual analysis—struggle to provide user-adaptive and communicatively effective explanations, especially in language-based applications [6, 7]. These limitations are especially concerning in public health and biomedical fields, where information must be not only technically sound but also effectively communicated to society. Recent work in human-computer interaction (HCI) has also emphasized the need for explainable, accountable, and intelligible systems [8]. Taken together, these challenges call for a new paradigm in explainability that

ArgXAI@ECAI: 3rd International Workshop on Argumentation for eXplainable AI, October 26, 2025, Bologna, Italy
*Corresponding author.
ilgenb@rki.de (B. İlgen); dubeya@rki.de (A. Dubey); hattabg@rki.de (G. Hattab)
0000-0001-5725-0850 (B. İlgen); 0009-0008-4823-9375 (A. Dubey); 0000-0003-4168-8254 (G. Hattab)

mirrors how humans reason and justify decisions. In this context, explanation ought to be understood as a reasoning process rather than merely a visualization or annotation.

Defining what constitutes an explanation is itself a complex issue. As reviewed in [9], explanations have been conceptualized in various ways: as assignments of causal responsibility [10], as both the process and product of addressing a "Why?" question [11], and as a means of constructing shared meaning. These perspectives highlight that explanation is not merely a factual output but a communicative and cognitive process that engages reasoning and interpretation. To this end, we propose PHAX: a Public Health Argumentation and eXplainability framework. PHAX is a multi-layer architecture integrating structured argumentation, adaptive natural language processing (NLP), and user modeling to generate clear, audience-specific justifications for AI outputs. It treats explanation not as a post-hoc add-on, but as a first-class component of decision-making pipelines. Structured argumentation functions as a core mechanism, allowing AI systems to explain their decision processes step by step, handle uncertainty, and reconcile conflicting evidence through formal reasoning [12]. Such capabilities are essential for building trust in AI-driven public health systems and biomedical systems. More specifically, within the domains of public health and biomedical sciences, we demonstrate how argumentation enhances explainability, with applications spanning areas such as decision-making (e.g., vaccination prioritization or clinical risk stratification), justification of system outputs (e.g., medical term simplification or the selection of diagnostic biomarkers), and interactive dialogue (e.g., clinician-AI interaction in diagnosis or treatment planning). These capabilities allow AI systems to deliver context-sensitive explanations aligned with stakeholder needs across both population-level and individual-level biomedical applications. Through structured reasoning and audience-aware communication, argumentation enables AI systems to provide transparent, tailored explanations across a range of high-stakes scenarios in public health and biomedical sciences.

PHAX builds on the formal tools of argumentation theory—including Dung's Abstract Framework [13] and ASPIC$^+$—to model [14] outputs as defeasible claims supported by reasoning chains. It also incorporates adaptive NLP techniques such as text simplification (TS), semantic role labeling (SRL), and discourse parsing, and audience-aware surface realization to tailor explanations to different users. Whether the audience is a patient, clinician, or policymaker, PHAX generates logically grounded and socially appropriate explanations.

To demonstrate the utility of PHAX, we present medical text simplification (MTS) as a core use case. Simplification decisions—such as replacing "myocardial infarction" with "heart attack" — are modeled as arguments, based on corpus frequency, semantic equivalence, and contextual appropriateness. Explanations are then adjusted in tone and depth based on user profiles. This showcases how PHAX enhances interpretability, transparency, and trust in a critical public health application. This paper makes the following contributions: **(i)** Introduces PHAX, a novel framework that integrates structured argumentation and adaptive NLP for explainable AI in public health and biomedical sciences., (ii) Demonstrates how simplification and other AI outputs can be modeled as defeasible reasoning chains., (iii) Proposes user-adaptive explanation strategies tailored to different stakeholders., (iv) Provides illustrative use cases highlighting PHAX's applicability in diverse public health contexts.

2. Related Work

XAI encompasses a range of approaches designed to make model behavior more interpretable. Common techniques include feature attribution methods (e.g., LIME, SHAP), saliency mapping, and counterfactual reasoning. These methods aim to provide insight into how AI models arrive at their predictions, but they often lack the ability to produce explanations that are user-adaptive and socially contextualized—particularly in domains like public health and biomedical sciences [15]. Given the shortcomings of purely statistical or post-hoc approaches, researchers have begun to investigate structured argumentation as a foundation for AI explanations. Frameworks based on Dung's Abstract Argumentation Framework (AF) and ASPIC$^+$ have been explored as mechanisms to model reasoning processes and support step-by-step justifications for AI outputs. For instance, Vassiliades et al.[12] and Čyras et al.[16]

survey a range of argumentation-based XAI approaches, showing how argument structures can provide more transparent and logically grounded explanations, particularly in settings involving uncertainty or conflicting information.

Although structured argumentation provides a strong basis for XAI, prior work has often emphasized symbolic and formal rigor over communicative usefulness, overlooking how explanations are interpreted by diverse users. PHAX builds on argumentation theory while extending it through user modeling and adaptive natural language generation to move beyond structural clarity. Unlike prior approaches, PHAX aims to deliver context-sensitive, stakeholder-specific justifications that are not only logically coherent but also socially meaningful.

Biomedical and healthcare research has provided concrete cases where argumentation-based explainability is applied. These studies indicate that argumentation theory helps clinicians reason under uncertainty and incomplete information. Longo et al. [17], for instance, applied defeasible reasoning and formal argumentation to model expert judgments in cancer recurrence prediction. This aligns with broader research in hybrid intelligence, which emphasizes collaborative, explainable AI systems that support human reasoning rather than replace it [18]. Seen in this light, argumentation and explanation are key elements in the design of transparent and reliable systems, a need that is especially pressing in healthcare contexts. Such approaches stress the importance of aligning machine reasoning with human cognitive and ethical expectations—an objective well supported by structured argumentation frameworks such as PHAX. By combining user-adaptive justifications with formal inference, PHAX supports this vision and helps ensure that explanations are not only logically sound but also socially meaningful in varied health contexts.

One concrete example of such a system is the CONSULT project [19], which applies computational argumentation to clinical settings. The CONSULT system brings together data from EHRs, wearable sensors, and treatment guidelines to aid collaborative decision-making. Using ASPIC^{+} to reason under uncertainty, it produces argumentation-based dialogues that explain treatment options to both patients and clinicians. By drawing on argument schemes, attack relations, and user-facing explanations, CONSULT mirrors the aims of PHAX in structuring and presenting personalized justifications for different stakeholders. Beyond such systems, recent work on user-adaptive explanation and NLG (e.g., [20]) highlights the importance of tailoring explanations to diverse audiences through role-sensitive or dialogue-based generation. However, these approaches are rarely integrated with structured argumentation, leaving a gap that PHAX directly addresses.

3. PHAX: Public Health Argumentation and eXplainability Framework

3.1. Architecture and Layers

PHAX (Public Health Argumentation and eXplainability) is a structured argumentation framework designed to enhance the transparency, accountability, and user alignment of AI systems in public health and biomedical domains. Embedding explainability into reasoning structures helps overcome the limits of post-hoc or model-agnostic XAI, especially in high-stakes domains. PHAX combines formal reasoning, NLP, and audience-aware methods to produce explanations that are both context-sensitive and socially meaningful. Unlike post-hoc methods, PHAX embeds explanation in the decision pipeline, allowing outputs to be justified and adapted to counterarguments and user needs. Table 1 shows how PHAX applies key XAI goals [12] through NLP tasks in public health. Its architecture has four layers, moving from raw data to user-adaptive explanation. As Figure 1 illustrates, each layer passes information forward, with user feedback enabling refinement. Argumentation serves as the core reasoning mechanism that translates NLP-derived insights into structured justifications. It connects the output of the NLP Layer to both the internal logic of decision-making and the external communicative needs of the user interface.

- **Data Layer:** The Data Layer gathers and preprocesses heterogeneous sources—such as clinical texts, patient records, epidemiological databases, and social media content—so that structured and unstructured inputs are harmonized before moving to subsequent layers.

- **NLP Processing Layer:** The NLP layer performs domain-specific analysis (NER, SRL, discourse parsing, text simplification), producing structured input for argument construction and validation.
- **Explanation and Argumentation Layer:** This layer models outputs as defeasible arguments with structures such as Dung's AF and ASPIC+. Arguments consist of claims (e.g., a simplification), supports (e.g., corpus frequency or semantic equivalence), and counterarguments (e.g., ambiguity or domain-specific issues). It formalizes reasoning and helps manage uncertainty and conflict.
- **User Interface Layer:** The user interface layer delivers explanations via dashboards, agents, or summaries, adapting tone and depth to the user (patient, clinician, policymaker). It links formal reasoning with human interpretability.

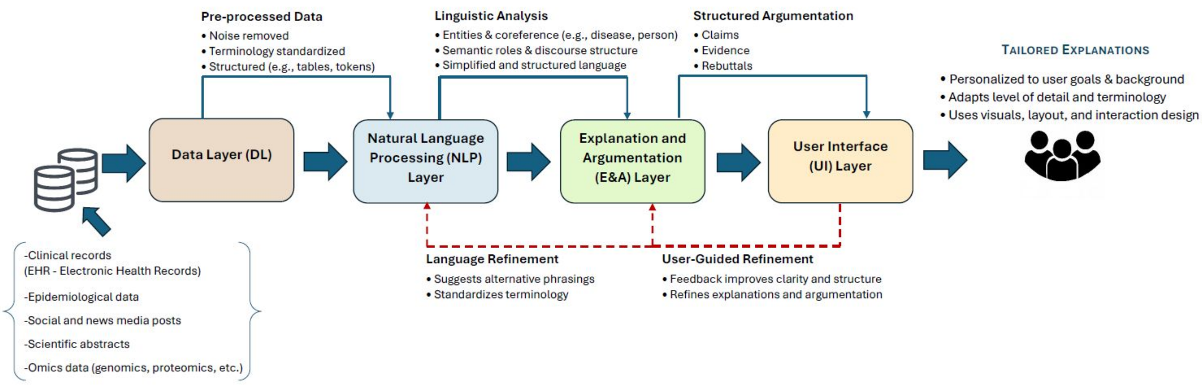

Figure 1: The PHAX layered architecture for user-centered explainable AI in public health.

Traditional frameworks separate development and deployment, but PHAX uses a layered design that integrates explanation processes across the AI lifecycle. Each layer supports model building and real-time explanation, ensuring traceability and stakeholder alignment.

Table 1
Mapping XAI Objectives to NLP Tasks in the PHAX Framework

XAI Objective	NLP Task	Illustrative Use Case
Transparency	Discourse Parsing, Semantic Role Labeling	Explaining vaccine recommendation steps in a logical sequence
Justification	Argumentation Mining, Natural Language Inference	Providing structured evidence for prioritizing vaccination groups
Relevance	Question Answering, Information Retrieval	Answering "Why are masks still needed?" with data-backed explanations
Conceptualization	Text Simplification, Named Entity Recognition	Simplifying terms like "PCR Test" for lay audiences
Learning	Dialogue Systems, Knowledge Graphs	Teaching how vaccines work through chatbot interactions

3.2. Formal Specification and Data Flow in PHAX

PHAX follows a layered architecture that integrates formal argumentation, NLP, and user modeling for user-adaptive explainability in public health. Each component is defined by its data types and transformation functions, enabling traceability from raw input to stakeholder-tailored explanations.

Formal Framework. We define an abstract argumentation framework as $AF = (A, R)$, where A is the set of arguments and $R \subseteq A \times A$ the attack relation. In PHAX/ASPIC^{+}, the knowledge base is $KB = (K_n, K_p)$, where K_n contains strict rules and facts, and K_p consists of defeasible rules and empirical observations.

Rule Types. Strict rules (R_s) encode deterministic knowledge (e.g., *clinical guideline* $\Rightarrow$ *recommendation*). Defeasible rules (R_d) capture uncertain, corpus- or context-driven inferences (e.g., *if frequency(symptom*$_A$*)* $>$ *frequency(symptom*$_B$*), prefer(symptom*$_A$*)*).

NLP Mapping. NLP modules (such as NER, SRL, and discourse parsing) map their outputs into premises and rules via:

$$\Phi : \text{NLP outputs} \rightarrow (P, R_s \cup R_d),$$

where P denotes extracted premises.

Argument Construction. Arguments are constructed by chaining premises and rules:

$$Arg = \langle P, Conclusion \rangle,$$

with attacks (rebut, undercut) following ASPIC^{+}. For clarity, we use this compact form, though ASPIC^{+} models arguments in full detail.

Graph Evaluation. Given semantics $\sigma \in \{\text{grounded}, \text{preferred}\}$, the accepted extension is $Ext_\sigma \subseteq A$. Grounded semantics yield conservative acceptance, while preferred allow richer extensions; other variants may also apply.

Explanation Object. A PHAX explanation is:

$$E = (T^*, U),$$

where T^* is the argument subtree supporting the output under σ, and U is the user profile. E is valid if T^* supports the claim and satisfies utility criteria such as readability, detail, or audience alignment. In other words, explanations are extracted from the evaluated argument graph (Ext_σ) and then tailored to the needs of the user profile U.

3.3. Common Argumentation Schemes in Public Health and Biomedical Reasoning

Decisions in public health and biomedicine often face uncertainty, multiple stakeholders, ethical issues, and context dependence. To handle these factors, argumentation may take different forms—causal, analogical, practical, or expert-based—depending on the task and audience. PHAX applies well-established argumentation schemes to generate explanations that are systematically organized and accessible to diverse users. These schemes capture typical reasoning patterns employed to justify claims in various domains [21]. Each scheme defines a type of inference (e.g., expert authority, practical goals) and is accompanied by critical questions that guide its evaluation. In public health and biomedical decisions, they provide a solid basis for user-facing justifications. Table 2 shows several schemes adapted to real-world scenarios. Beyond structuring logical support, they also act as templates for natural language explanations aligned with stakeholder needs.

Formal Representation of Schemes

PHAX uses formal representations of argumentation schemes to support logic-based justification and reasoning. Below are selected examples:

Table 2
Argumentation Schemes (AS) in Public Health and Biomedical Reasoning

AS	Description	Example in Context
Expert Opinion	Relying on authority or professional expertise	"WHO recommends vaccination for this age group"
Cause to Effect	Predicting consequences of an action or event	"Masking reduces viral transmission"
Practical Reasoning	Choosing actions to achieve desired outcomes	"To prevent ICU overload, implement lockdown"
Analogy	Inferring based on similarity to previous cases	"Contact tracing worked for Ebola; it can help for COVID"
Statistical Generalization	Drawing conclusions from population-level data	"This drug helped 70% of patients in clinical trials"
Ethical/Value-based	Arguing based on fairness, harm, or social values	"We must prioritize vulnerable groups to ensure equity"

Formalization: The scheme for $ExpertOpinion$ where P is the proposition under consideration and D is the relevant domain, can be encoded as:

$$\text{is\_expert}(E, D),\ \text{asserts}(E, P),\ \text{relevant}(P, D) \Rightarrow \text{believe}(P)$$

Cause to Effect:

$$\text{action}(A),\ \text{causes}(A, E) \Rightarrow \text{expect}(E)$$

Practical Reasoning:

$$\text{goal}(G),\ \text{action}(A),\ \text{promotes}(A, G) \Rightarrow \text{do}(A)$$

3.4. Structured Reasoning and Argumentative Explanation in PHAX

To support structured and adaptable explanations, PHAX relies on a hybrid formal foundation that combines elements from deductive, structured, and label-based argumentation models. PHAX applies structured argumentation to ensure that conclusions remain traceable and valid in biomedical and public health contexts. The framework incorporates ASPIC+, which supports both strict (deductive) and defeasible rules. Strict rules model clear-cut logic (e.g., eligibility based on clinical criteria), while defeasible rules capture reasoning under uncertainty and exceptions—crucial in high-stakes public health decision-making.

Additionally, PHAX adopts principles from label-based argumentation to handle preference, uncertainty, and credibility. Arguments may carry labels such as confidence, stakeholder relevance, or ethical weight; these propagate through the argument graph to guide resolution. This makes the system sensitive to contextual and user-specific needs, supporting more personalized and socially attuned justifications. Finally, PHAX incorporates argumentation schemes, such as Expert Opinion, Practical Reasoning, and Cause to Effect, which reflect common patterns of human reasoning. These schemes serve as templates for generating natural language explanations that align with how different stakeholders interpret justification—enhancing both transparency and persuasive power.

At the core of PHAX is the use of structured argumentation to represent and explain AI-generated outputs. Each decision is modeled as a claim supported by explicit premises and, when appropriate, challenged by potential objections—mirroring human reasoning and enabling transparent justifications. For instance, in medical text simplification, PHAX treats the decision to simplify a term X to Y not merely as an output, but as an argument that can be analyzed and, if needed, contested.

3.4.1. Illustrative Example: Medical Simplification as Structured Argument

The following structures are formalized using ASPIC$^+$, enabling both graphical visualization and logic-based evaluation. This approach goes beyond surface-level explainability by exposing the reasoning process itself. In particular, the decision to simplify a term X to Y is not presented as a final output alone, but accompanied by explicit justifications and possible objections. This aligns with the principles of structured argumentation used in explainable AI.

Claim	Term X is simplified to Y
Support 1	Y is more frequent in lay corpora
Support 2	No semantic loss detected via NLI model
Attack	Y may be ambiguous in clinical contexts

Dung's Abstract Argumentation Framework (AF)

In Dung's AF [13], arguments are modeled as atomic elements with defined attack relations. Let:

A: The argument supporting the simplification of $X \to Y$
B: Support based on frequency: "Y is more frequent than X"
C: Support based on semantic similarity: "No meaning lost"
D: Counterargument: "Y is ambiguous in clinical contexts"

We define the argument set and the attack relation as follows:

$$\text{Args} = \{A, B, C, D\}, \quad \text{Attacks} = \{(D, A)\}$$

Here, D challenges the simplification decision, which can be evaluated using grounded or preferred semantics depending on the context.

ASPIC$^+$ Representation

ASPIC$^+$ [14] enriches this view by including internal structure, rules, and types of reasoning. The same example can be modeled as:

Premises

P_1: frequency$(Y) >$ frequency(X)
P_2: semantic_match$(X, Y) = True$
P_3: ambiguity$(Y) = High_Clinical$

Rules

$r_1 : (P_1, P_2) \Rightarrow prefer(Y)$
$r_2 : P_3 \Rightarrow \neg prefer(Y)$

Arguments

$Arg_1 = \langle P_1, P_2, r_1 \rangle \Rightarrow prefer(Y)$
$Arg_2 = \langle P_3, r_2 \rangle \Rightarrow \neg prefer(Y)$

Here, Arg_2 attacks Arg_1, resulting in a defeasible justification structure. The system can select or reject the simplification based on external preferences, such as the user's role (e.g., patient or clinician). This structured approach allows PHAX to generate explainable outputs that go beyond readability scores, instead providing reasoned justifications that can be tailored and interrogated across use cases.

3.4.2. Formalization: Evidence-Based Reasoning via PICO

Moving beyond basic linguistic tasks such as term simplification, PHAX's formal reasoning capabilities extend to evidence-based clinical logic. The following illustrates how structured argumentation can be applied to biomedical literature analysis, using the widely adopted PICO paradigm. Building on the earlier simplification use case, we now illustrate how the same argumentation machinery can support evidence-based clinical reasoning through the PICO paradigm. In particular, structured argumentation offers a compelling foundation for modeling *evidence-based claims* derived from biomedical literature using the PICO (Population, Intervention, Comparison, Outcome) paradigm. PICO elements can be expressed as formal predicates, enabling the construction of defeasible rules (see below).

Predicates
- $P(x)$: entity x belongs to target population
- $I(x)$: intervention applied
- $C(x)$: control/comparison condition
- $O(x)$: observed or expected outcome

Defeasible Rule

$$P(x) \wedge I(x) \Rightarrow_O O(x)$$

This implies that for individuals in population P, the application of intervention I leads to outcome O—under typical conditions. However, due to co-morbidities, alternative studies, or contextual constraints, such a rule remains *defeasible*. Counterarguments may cite exceptions (e.g., "I is contraindicated for subgroups in P"). Using ASPIC$^+$, such clinical evidence can be formalized as follows:

Premises

P_1: Study population matches P
P_2: Intervention I applied
P_3: Outcome O observed
P_4: Source study credible

Defeasible Rule

$$(P_1, P_2, P_3, P_4) \Rightarrow \text{recommend}(I, P)$$

Counterargument

$$(P_1', P_2', \neg O, P_4') \Rightarrow \neg\text{recommend}(I, P)$$

These opposing arguments can then be compared via preference criteria (e.g., study quality, sample size) and evaluated within an argumentation framework using grounded or preferred semantics. Grounded semantics selects the most cautious acceptable set of arguments, while preferred semantics favors maximal admissible sets. This formalism not only enhances the interpretability of AI recommendations in public health contexts, but also allows systematic traceability of how and why a certain intervention is proposed—bridging evidence-based medicine and explainable AI.

3.5. User-Adaptive Explanation Generation

Public health communication involves a range of stakeholders—clinicians, policymakers, patients—each with different cognitive needs and expectations. To support effective communication, PHAX dynamically adapts both the structure and the presentation of its explanations based on user profiles. These user modeling attributes—such as expertise, lexical tolerance, and cognitive expectations—govern how explanations are tailored across multiple adaptation layers, as illustrated below.

Theoretical Foundation. Drawing from Relevance Theory [22] and Grice's Cooperative Principles [23], PHAX ensures that explanations are not only accurate but also cognitively appropriate for the intended audience. This is operationalized through user modeling and selective generation of explanation content. Furthermore, PHAX incorporates principles from Labelled Argumentation Frameworks [24, 25] to propagate metadata such as confidence, role-based preference, or ethical weight across the explanation graph.

Definition 1. (User Profile) A user profile U is a tuple (e, l, c), where:
- $e \in \mathbb{R}$: Domain expertise level (e.g., clinician vs. layperson)
- $l \in \mathbb{R}$: Lexical tolerance (e.g., jargon sensitivity)
- $c \in \mathbb{R}$: Cognitive depth (e.g., expected explanation complexity)

Definition 2. (Semantic Sufficiency) Given an explanation tree T and argument a, semantic sufficiency $\sigma_T(a) \in [0, 1]$ quantifies the extent to which T supports a, possibly via aggregation over leaf node support and edge weights.

Definition 3. (Utility Function) Utility is a linear combination of weighted factors:

$$\text{Utility}(T, U) = \sum_{i=1}^{n} w_i \cdot f_i(T, U)$$

where f_i is a feature function (e.g., clarity, lexical fit), and $w_i \in \mathbb{R}$ is a tunable weight.

Formal Mechanism. Each user is modeled as a profile U with attributes defined in the *user profile*. Given a full Quantitative Dispute Tree $QDT(a)$ [26] for an argument a, the framework selects a user-appropriate subgraph T^* as follows:

$$T^* = \arg\max_{T} \text{Utility}(T, U) \quad \text{subject to} \quad \sigma_T(a) \geq \tau(a)$$
$$\text{Utility}(T, U) = \alpha \cdot \text{Clarity}(T, U) + \beta \cdot \text{Relevance}(T, U) + \gamma \cdot \text{LexicalFit}(T, U)$$

Where:
- $\sigma_T(a)$: Semantic sufficiency — does T still justify argument a?
- $\tau(a)$: Task-defined threshold for completeness

Adaptation Dimensions. Adaptation operates along lexical complexity (simplified phrasing for lay users), information depth (detailed chains for experts, summaries for general audiences), and presentation format (visuals for policymakers, text for patients, dialogue for clinicians).

Illustrative Example. A vaccine prioritization decision may be explained as **clinical evidence for a clinician** ("Phase III trial data show 92% efficacy"), **personal reassurance for a patient** ("This vaccine has helped many people like you stay safe"), or **system-level impact for a policymaker** ("Prioritizing this group prevents ICU overload by 45%").

Relation to Argumentation Schemes. User-tailored explanations map to different argumentation schemes depending on the audience: *Cause to Effect* for lay users ("Vaccination reduces risk of severe disease"), *Statistical Generalization* for experts ("70% of patients showed improvement"), *Practical Reasoning* for decision-makers ("To prevent ICU overload, prioritize group A"), and *Ethical Reasoning* for public discourse ("We must protect the most vulnerable first").

Connection to User Interface Layer. These adaptive explanations are operationalized through the User Interface Layer of PHAX, which selects and renders the appropriate format and depth of explanation based on the computed utility for each user profile. The UI layer delivers argument structures in different formats—textual justifications, interactive dialogues, or visual dashboards—acting as the channel through which they reach the user. In this way, the formal reasoning developed in earlier layers is preserved while being presented in a form that is both understandable and convincing for its audience.

4. Application Scenarios Across Public Health and Biomedical Sciences

PHAX addresses a broad spectrum of reasoning and communication challenges in public health and biomedical domains, where decisions often involve uncertainty, competing values, and diverse stakeholders. Beyond its core architecture, the framework provides structured and audience-sensitive explanations tailored to real-world needs—from clinical decision support to public communication. Below, we present representative scenarios illustrating how PHAX integrates argumentation and explanation to promote transparency, trust, and actionable insight across practical settings.

4.1. Decision Support and Stakeholder Alignment

Public health decisions frequently demand balancing competing priorities, working with limited resources, and dealing with uncertainty. For instance, setting vaccination priorities during a pandemic requires weighing exposure risks, equity concerns, and the capacity of the healthcare system. PHAX models such dilemmas using defeasible argumentation, enabling transparent, traceable justifications for complex decisions. Its layered architecture delivers explanations tailored to different stakeholders: clinicians may explore structured evidence trails via interactive dashboards, while policymakers access high-level summaries that emphasize societal trade-offs and ethical considerations.

4.2. Evidence Synthesis and Biomedical Summarization

Systematic reviews play a central role in biomedical research by combining results from multiple studies, but their length and variability can make them hard to access and interpret. PHAX supports structured summarization by using argumentation mining on PICO-extracted data to capture key claims, counterclaims, and the strength of supporting evidence. These elements are organized into argument structures, producing contrastive summaries that highlight where studies agree, disagree, or remain uncertain. Such summaries help clinicians and researchers quickly navigate complex and often conflicting literature.

4.3. Public Communication and Policy Justification

Effective communication of health interventions—such as lockdowns or vaccine mandates—requires balancing scientific accuracy with accessibility for diverse audiences. PHAX supports this by using established argumentation schemes (causal, ethical, practical) and adapting their wording and framing to different user profiles. For instance, a lockdown policy may be framed in terms of "transmission control" when addressing clinicians, but emphasized as "protecting the vulnerable" in public-facing messages. This audience-sensitive adaptation enhances clarity and trust without compromising factual integrity.

4.4. Risk Communication and Misinformation Rebuttals

Health misinformation often spreads through arguments that are emotionally compelling but logically weak. PHAX tackles this by producing structured rebuttals: it breaks claims into premises, tests their validity, and formulates counterarguments supported by scientific evidence and adapted to the audience. For instance, the false claim that "vaccines cause infertility" can be refuted through mechanistic evidence and trial data for clinicians, while lay audiences may receive simpler, empathetically framed responses that emphasize safety and social consensus. This audience-aware rebuttal strategy enhances persuasive effectiveness without compromising scientific rigor.

4.5. Interface-Driven Personalization and Delivery

An explanation is shaped as much by how it is delivered as by what it contains. PHAX addresses this by offering different modes of presentation, tailored to user preferences, literacy level, and context. These modes include narrative text, visual overviews, and conversational dialogue. For example, patients may receive conversational explanations via chatbot interfaces, while policymakers might explore comparative scenario graphs that highlight trade-offs. These modalities are selected dynamically based on user modeling, ensuring that the explanation aligns with the user's cognitive and informational needs as captured by PHAX's adaptive layer.

5. Conclusion and Future Work

This study presents PHAX—a Public Health Argumentation and eXplainability framework—designed to support transparent, context-aware, and user-adaptive explanations in high-stakes domains such as healthcare and biomedical sciences. Building upon structured argumentation theory, PHAX incorporates formal reasoning, adaptive NLP pipelines, and user modeling to generate stakeholder-specific justifications for AI outputs. Our contributions include a modular architecture for integrating explainability into the AI lifecycle, a formalization of user-adaptive explanation generation, and illustrative applications in medical term simplification, policy justification, and systematic review summarization. By combining defeasible reasoning, argumentation schemes, and multimodal delivery interfaces, PHAX enables interpretable decision support tailored to diverse user needs. Future work extends PHAX with uncertainty-aware and value-sensitive argumentation to better reflect complex, conflicting public health priorities. A key direction involves activating PHAX's adaptive layer through live user feedback, enabling continuous refinement of explanations aligned with user profiles. We also plan to evaluate PHAX in real-world settings through user studies with clinicians, policymakers, and patients, to assess explanation effectiveness, trust calibration, and usability in practice.

Acknowledgments

This work has been financially supported by the German Federal Ministry of Health (BMG) under grant No.: ZMI5- 2523GHP027 (project "Strengthening National Immunization Technical Advisory Groups and their Evidence-based Decision-making in the WHO European Region and Globally" SENSE) part of the Global Health Protection Programme, GHPP.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] E. J. Topol, High-performance medicine: The convergence of human and artificial intelligence, Nature Medicine 25 (2019) 44–56. doi:10.1038/s41591-018-0300-7.

[2] J. Amann, A. Blasimme, E. Vayena, D. Frey, V. I. Madai, P. Consortium, Explainability for artificial intelligence in healthcare: A multidisciplinary perspective, BMC Medical Informatics and Decision Making 20 (2020) 310. doi:10.1186/s12911-020-01332-6.

[3] G. Hattab, C. Irrgang, N. Körber, D. Kühnert, K. Ladewig, The way forward to embrace artificial intelligence in public health, American Journal of Public Health 115 (2025) 123–128. doi:10.2105/AJPH.2024.307888.

[4] A. Dubey, Z. Yang, G. Hattab, A nested model for ai design and validation, iScience 27 (2024) 110603. doi:10.1016/j.isci.2024.110603.

[5] H. Khalili, M. A. Wimmer, Towards improved xai-based epidemiological research into the next potential pandemic, Life 14 (2024) 783. doi:10.3390/life14070783.

[6] D. Mindlin, A. S. Robrecht, M. Morasch, P. Cimiano, Measuring user understanding in dialogue-based xai systems, in: ECAI 2024 - 27th European Conference on Artificial Intelligence, 19-24 October 2024, Santiago de Compostela, Spain - Including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024), 2024, pp. 1148–1155. doi:10.3233/FAIA240608.

[7] G. He, N. Aishwarya, U. Gadiraju, Is conversational xai all you need? human-ai decision making with a conversational xai assistant, in: Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI 2025), ACM, 2025, pp. 907–924. doi:/10.1145/3708359.3712133.

[8] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, M. Kankanhalli, Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda, in: Proceedings of

the 2018 CHI Conference on Human Factors in Computing Systems, ACM, 2018, pp. 1–18. doi:10.1145/3173574.3174156.
[9] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence 267 (2019) 1–38. doi:/10.1016/j.artint.2018.07.007.
[10] J. R. Josephson, S. G. Josephson, Abductive Inference: Computation, Philosophy, Technology, Cambridge University Press, 1996.
[11] T. Lombrozo, The structure and function of explanations, Trends in Cognitive Sciences 10 (2006) 464–470. doi:10.1016/j.tics.2006.08.004.
[12] A. Vassiliades, N. Bassiliades, T. Patkos, Argumentation and explainable artificial intelligence: a survey, The Knowledge Engineering Review 36 (2021) e5. doi:/10.1017/S0269888921000011.
[13] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artificial Intelligence 77 (1995) 321–357. doi:10.1016/0004-3702(94)00041-X.
[14] S. Modgil, H. Prakken, A general account of argumentation with preferences, Artificial Intelligence 195 (2013) 361–397. doi:10.1016/j.artint.2012.10.008.
[15] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, R. Ranjan, Explainable ai (xai): Core ideas, techniques, and solutions, ACM Computing Surveys 55 (2023) 1–33. doi:doi.org/10.1145/3561048.
[16] K. Čyras, A. Rago, E. Albini, P. Baroni, F. Toni, Argumentative xai: A survey, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21), 2021, pp. 4392–4399. doi:10.24963/ijcai.2021/600.
[17] L. Longo, B. Kane, L. Hederman, Argumentation theory in health care, in: Proceedings of the 25th IEEE Symposium on Computer-Based Medical Systems, 2012. doi:10.1109/CBMS.2012.6266323.
[18] Z. Akata, D. Balliet, M. de Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos, H. Hung, C. Jonker, C. Monz, M. Neerincx, F. Oliehoek, H. Prakken, S. Schlobach, L. van der Gaag, F. van Harmelen, H. van Hoof, B. van Riemsdijk, A. van Wynsberghe, R. Verbrugge, B. Verheij, P. Vossen, M. Welling, A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence, Computer 53 (2020) 18–28. doi:10.1109/MC.2020.2996587.
[19] N. Kökciyan, M. Chapman, P. Balatsoukas, I. Sassoon, K. Essers, M. Ashworth, V. Curcin, S. Modgil, S. Parsons, E. Sklar, A collaborative decision support tool for managing chronic conditions, in: MEDINFO 2019: Health and Wellbeing e-Networks for All, IOS Press, 2019, pp. 644–648. doi:10.3233/SHTI190302.
[20] J. Delaunay, C. Largouët, L. Galárraga, N. Van Berkel, Adaptation of ai explanations to users' roles, in: Proceedings of the Workshop on Human-Centered Explainable AI (HCXAI 2023) at CHI, Hamburg, Germany, 2023, pp. 1–7. URL: https://nielsvanberkel.com/files/publications/chi2023f.pdf.
[21] D. Walton, C. Reed, F. Macagno, Argumentation Schemes, Cambridge University Press, Cambridge, 2008.
[22] D. Sperber, D. Wilson, Relevance: Communication and Cognition, volume 142, Harvard University Press, Cambridge, MA, 1986.
[23] H. P. Grice, Logic and conversation, in: P. Cole, J. L. Morgan (Eds.), Syntax and Semantics: Volume 3 - Speech Acts, Academic Press, 1975, pp. 41–58.
[24] L. Amgoud, J. Ben-Naim, Ranking-based semantics for argumentation frameworks, in: Proceedings of the International Conference on Scalable Uncertainty Management (SUM), Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 134–147. doi:10.1007/978-3-642-40381-1_11.
[25] A. Hunter, A probabilistic approach to modelling uncertain logical arguments, International Journal of Approximate Reasoning 54 (2013) 47–81. doi:/10.1016/j.ijar.2012.08.003.
[26] K. Čyras, T. Kampik, Q. Weng, Dispute trees as explanations in quantitative (bipolar) argumentation, in: Proceedings of the 1st International Workshop on Argumentation for Explainable AI (ArgXAI), volume 3209, CEUR Workshop Proceedings, 2022, pp. 1–12. URL: https://ceur-ws.org/Vol-3209/0872.pdf.