Privacy Firewall for Older Adults: A Semantic Gateway for Trustworthy Ambient Assisted Living

Tim Markwardt[1][0000-0002-7577-5519],
Gerald Bieber[1][0000-0003-2496-6232],
Purbaditya Bhattacharya[2][0000-0003-1679-0900], and
Christian Peter[3][0009-0002-7423-6852]

[1] Fraunhofer-Institut für Graphische Datenverarbeitung IGD, Rostock, Germany
tim.markwardt, gerald.bieber@igd-r.fraunhofer.de
[2] University of Rostock, Rostock, Germany
purbaditya.bhattacharya@uni-rostock.de
[3] tech@spree GmbH, Rostock, Germany christian.peter@spree.de

Abstract. Ambient Assisted Living (AAL) systems continuously collect highly sensitive data within the homes of older adults, a group that often cannot fully assess the privacy implications of sharing such information. Conventional safeguards protect data in transit and at rest, and privacy-preserving AI governs where and how data are processed, yet none continuously evaluate *what* multimodal information should leave the home at all. We introduce the concept of a *Personal AI Firewall*, a semantic privacy gateway that evaluates the content and context of outgoing information entirely on local edge hardware and selectively blocks, redacts, anonymises, aggregates or abstracts it according to context-aware policies. We derive five design principles from the specific needs of older adults and present a layered edge architecture comprising data ingestion, semantic analysis, a context-aware privacy policy engine, data transformation and an explaining output gateway, all of which keep raw data on-premises. We further propose a structured evaluation methodology spanning technical feasibility, privacy effectiveness and user acceptance, and critically discuss the paradigm's central open challenges. By reframing the firewall from a network construct into an intelligent, content-aware privacy intermediary, this position paper contributes a new paradigm for trustworthy, human-centred data governance in ageing societies.

Keywords: Personal AI firewall, Ambient Assisted Living (AAL), Privacy-preserving AI, Edge AI, Older adults, Semantic gateway

1 Introduction

The increasing deployment of sensor-based Ambient Assisted Living (AAL) technologies, smart home devices, video communication systems, and artificial intel-

ligence offers significant opportunities to support older adults in maintaining independence, safety, and quality of life [11,66]. However, these technologies inevitably collect large amounts of sensitive personal data within the private home environment. Many older adults possess limited digital literacy and are often unable to fully assess the privacy implications of sharing personal information with external services, healthcare providers, or cloud-based AI systems [29,50]. Consequently, existing technical solutions frequently expose users to unintended privacy risks despite complying with conventional cybersecurity measures [48,35,53,31].

This paper introduces the concept of a Personal AI Firewall, a novel privacy-preserving middleware that operates as a semantic protection layer between the domestic environment and external digital services. Unlike traditional network firewalls, which regulate communication based on network protocols, ports, and access rules, the proposed system evaluates the content and context of outgoing information using artificial intelligence. Multimodal data originating from cameras, environmental sensors, wearable devices, microphones, and communication platforms are continuously analysed to identify privacy-sensitive content. Based on configurable privacy policies, user preferences, and contextual reasoning, the Personal AI Firewall selectively blocks, redacts, anonymises, aggregates or abstracts information before it leaves the trusted home environment.

The proposed architecture aims to empower older adults by providing automated privacy protection without requiring technical expertise or active user intervention. It establishes a trustworthy intermediary that balances the benefits of digital health monitoring and remote assistance with the fundamental right to informational self-determination. Furthermore, the concept supports transparent and explainable decision-making, allowing caregivers, healthcare professionals, and users to understand why specific information is transmitted or withheld.

This paper is deliberately positioned as a concept and position paper: its aim is to define and motivate a new architectural paradigm rather than to report a completed empirical study. Concretely, our contributions are:

1. We introduce the concept of a Personal AI Firewall, a semantic privacy gateway that shifts privacy protection in AAL from governing *where* data may flow (network-level security) and *how* it is computed (process-level, privacy-preserving AI) towards continuously evaluating *what* multimodal information should leave the home, independent of its originating device or service.
2. We derive five design principles for such a system from the specific needs of older adults, and we present a layered edge architecture: data ingestion, semantic analysis, context-aware privacy policy engine, data transformation, and an explaining output gateway, that realises them while keeping all raw data on-premises.
3. We propose a structured evaluation methodology spanning technical feasibility, privacy effectiveness, and user acceptance, together with a set of public benchmark datasets and metrics, to guide the empirical validation of a future prototype.

4. We critically discuss the central open challenges of the paradigm: the break-glass safety dilemma, edge-resource limits, the reliability of AI-based decisions, and questions of accountability, to provide a research agenda for the community.

By reframing the firewall from a network construct into an intelligent, content-aware privacy intermediary, the Personal AI Firewall contributes a new paradigm for trustworthy, human-centred AI in ageing societies.

2 Related work

2.1 Ambient Assisted Living and Smart Home Technologies

The demographic shift towards an ageing society has spurred growing research activities in the field of Ambient Assisted Living (AAL) [11]. AAL systems aim to support older adults in maintaining independence, safety, and quality of life by integrating intelligent technologies into the domestic environment. Typical AAL deployments combine environmental sensors, wearable devices, smart appliances, cameras, microphones, and communication platforms to continuously monitor activities of daily living, detect emergencies, and facilitate remote care [5,11,66].

Recent advances in Internet of Things (IoT) technologies and artificial intelligence have substantially improved the capabilities of such systems. Machine learning algorithms can recognise daily routines, detect abnormal behaviour, predict health risks, and assist caregivers through automated analysis of heterogeneous sensor data [10,66]. Cloud computing further enables scalable storage and processing of multimodal information, allowing increasingly sophisticated healthcare services.

Despite these technological advances, the widespread adoption of AAL systems remains limited [10,50]. One of the primary barriers identified in numerous studies is the concern about privacy and the continuous collection of highly sensitive personal information inside the private home [50,3,51].

2.2 Privacy Challenges in Intelligent Home Environments

Unlike conventional IoT applications, AAL systems collect information that directly reflects an individual's health status, daily routines, social interactions, and living conditions. Video streams, audio recordings, motion trajectories, physiological measurements, and behavioural patterns together create highly detailed user profiles.

Although existing systems generally implement standard cybersecurity mechanisms such as encrypted communication, authentication, access control, and secure cloud infrastructures, these measures primarily protect data during transmission and storage [53]. They do not address whether the transmitted information should leave the private environment in the first place [35,48].

This distinction is particularly relevant for older adults. Many older adults possess limited digital literacy and may not fully understand the consequences of

granting extensive data access to external services [29,50]. Furthermore, consent is often obtained only once during system setup, while the sensitivity of collected information continuously changes depending on context [44,21].

Consequently, current privacy mechanisms largely rely on organisational policies, legal regulations such as the General Data Protection Regulation (GDPR) [16], and trust in service providers rather than technical mechanisms that actively minimise data disclosure.

2.3 Privacy-Preserving Artificial Intelligence

To address these concerns, considerable research has focused on privacy-preserving AI. Federated Learning enables machine learning without transferring raw training data to centralised servers [28], while Differential Privacy protects statistical information by introducing carefully calibrated noise [1,61,19]. Homomorphic Encryption and Secure Multi-Party Computation allow computations on encrypted data, although they remain computationally demanding [2,32] for many real-time AAL applications.

More recently, Edge AI has emerged as a promising approach. By performing inference directly on local computing devices, sensitive information can remain within the user's home, thereby reducing communication with cloud infrastructures. Modern edge platforms are increasingly capable of executing deep neural networks for image recognition, speech processing, and activity recognition with acceptable latency and energy consumption [40,37].

While these approaches significantly reduce privacy risks, they primarily focus on where data are processed rather than which information should be disclosed [1,28,40]. Most systems still transmit inference results, metadata, or selected sensor streams without considering their semantic privacy implications.

2.4 Context-Aware Privacy Protection

Several recent research projects have proposed context-aware privacy mechanisms that dynamically decide what information may be shared based on the task, the actors involved, and the user's predefined privacy preferences, often grounded in the framework of contextual integrity [8,38,20]. Similarly, explainable AI techniques seek to increase transparency by providing understandable justifications for algorithmic decisions [49,4].

Computer vision research has also investigated privacy filters capable of blurring faces and bodies, removing backgrounds, or replacing individuals with abstract avatars [65,24,54,30]. Other approaches anonymise speech or suppress identifiable information before transmission [22,58].

Although these techniques contribute important building blocks, they are typically designed for individual data modalities or isolated application scenarios. They rarely combine heterogeneous sensor information into a unified privacy assessment, nor do they continuously evaluate whether outgoing information complies with individual privacy expectations.

2.5 Research Gap

The literature demonstrates substantial progress in secure communication, privacy-preserving machine learning, edge computing, and context-aware access control. Nevertheless, existing approaches generally treat privacy protection as a collection of isolated technical mechanisms rather than as a continuous semantic decision process [39,51].

To the best of our knowledge, there is currently no comprehensive architecture that functions as an intelligent intermediary between the private home and external digital services by semantically evaluating all outgoing information independent of its source. While preliminary frameworks have emerged that leverage Large Language Models to configure user privacy preferences over sensor data flows [59], they presuppose a cooperative sensing infrastructure, operate over predefined sensor streams within a single environment, and are evaluated on isolated tasks rather than acting as a continuous, source-independent semantic gateway for *all* outgoing information. Similarly, network-layer approaches to controlling information exposure deliberately restrict themselves to traffic *destinations* and explicitly treat payload interception and content inference as out of scope [35,48], while privacy-preserving AI methods typically protect the computational process rather than evaluate the semantic content of transmitted data [28,40]. Consequently, existing systems address either *where* data flows or *how* it is computed, but none continuously evaluate *what* multimodal information leaves the trusted environment, independent of its originating device or service.

This paper addresses this gap by introducing the concept of a Personal AI Firewall, an AI-based semantic privacy gateway that continuously analyses multimodal information generated within the home environment. Instead of merely protecting communication channels, the proposed system determines whether outgoing information is privacy compliant, contextually appropriate, and necessary for the intended service before any data leave the trusted environment. In doing so, it extends the traditional concept of a firewall from network security towards intelligent, content-aware privacy protection tailored to the specific needs of older adults.

3 Concept and Methodological Framework

3.1 Design Principles

The Personal AI Firewall is guided by five core design principles derived from the requirements of older adults in AAL environments:

1. **Privacy by Default.** All outgoing information is assumed privacy-sensitive unless explicitly classified as safe by the system. No data leave the trusted home environment without semantic evaluation.
2. **Minimal Disclosure.** Only the minimum amount of information necessary to fulfil the intended service purpose is transmitted. Wherever possible, raw data are replaced by aggregated summaries, abstracted representations, or anonymised derivatives.

3. **Autonomy Without Complexity.** The system must provide meaningful privacy protection without requiring technical expertise or continuous active decision-making from the user. Default profiles and adaptive learning reduce the configuration burden on older adults.
4. **Transparency and Explainability.** Every filtering or blocking decision must be traceable and explainable in simple, non-technical language, enabling users, caregivers, and healthcare professionals to understand and adjust the system's behaviour.
5. **Contextual Adaptivity.** Privacy decisions are not static but depend on the current situation, including the user's activity, health status, communication partner, time of day, and urgency level.

3.2 System Architecture Overview

The proposed architecture is structured as a layered middleware positioned between the domestic sensor and communication infrastructure and all external digital services. Figure 1 illustrates the high-level architecture, which comprises five functional layers:

Table 1: Functional layers of the Personal AI Firewall architecture.

Layer	**Function**	**Examples**
Data Ingestion Layer	Captures and normalises multimodal data streams	Camera feeds, microphone input, wearable telemetry, environmental sensors
Semantic Analysis Layer	Interprets content, context, and sensitivity of information	Object/person detection, activity recognition, speech understanding
Privacy Policy Engine	Evaluates whether information complies with privacy rules	Rule matching, contextual reasoning, risk scoring
Data Transformation Layer	Applies appropriate privacy-preserving operations	Blocking, redaction, anonymisation, aggregation, abstraction
Output Gateway	Controls final transmission to external services	Selective forwarding, logging, explanation generation

All layers operate locally on edge computing hardware within the user's home. Only information that has passed through all layers and has been classified as permissible is forwarded to external recipients.

3.3 Data Ingestion and Multimodal Fusion

The Data Ingestion Layer interfaces with the heterogeneous sensor ecosystem typically found in AAL environments. It receives continuous streams from:

– **Visual sensors** (RGB cameras, depth sensors),

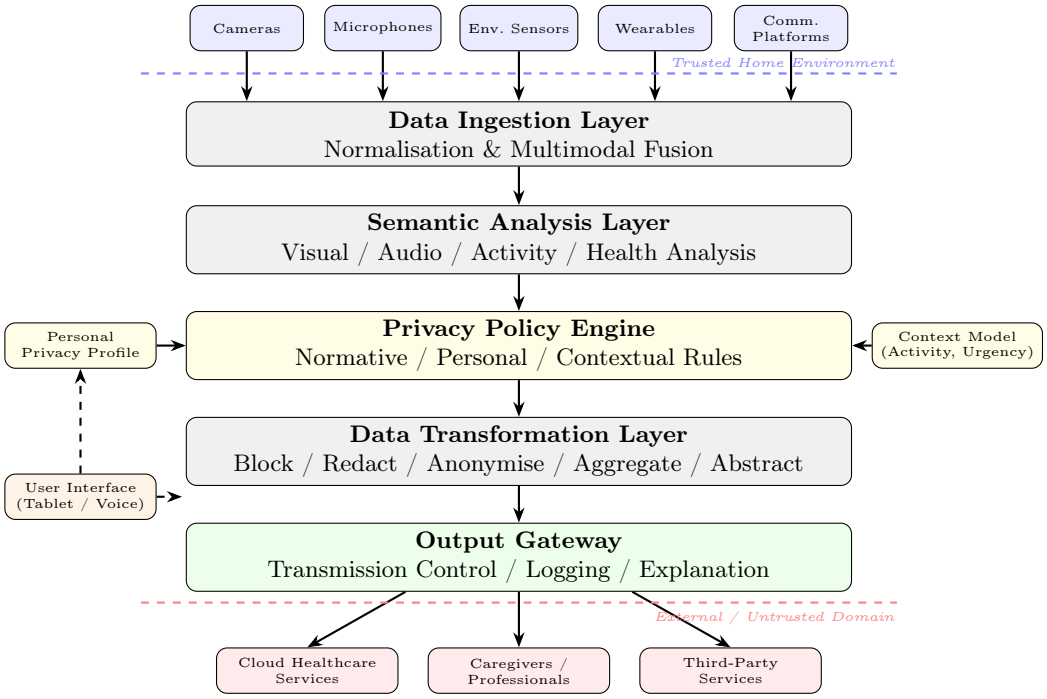

Fig. 1: High-level architecture of the Personal AI Firewall. All processing occurs within the trusted home environment. Only semantically evaluated and transformed data pass through the Output Gateway to external services.

- **Audio sensors** (microphones for voice interaction and ambient sound),
- **Environmental sensors** (temperature, humidity, door/window contacts, motion detectors),
- **Wearable devices** (heart rate monitors, accelerometers, pulse oximeters),
- **Communication platforms** (video calls, messaging applications, telehealth interfaces).

Each data stream is normalised into a unified internal representation that preserves temporal alignment and source metadata. A multimodal fusion module combines information across modalities to construct a coherent situational model of the home environment, enabling the subsequent layers to reason about context rather than isolated sensor readings.

3.4 Semantic Analysis Layer

The Semantic Analysis Layer constitutes the core intelligence of the Personal AI Firewall. It employs a set of specialised deep learning models operating on the edge device to extract semantic meaning from raw sensor data. Naturally, the

scope of semantic understanding is bounded by what sensor data is available in the given AAL environment:

- **Visual Analysis Module.** Performs person detection, face recognition, pose estimation, and scene understanding. Identifies whether privacy-sensitive content is present, such as recognisable faces, exposed bodies, medical devices, or documents. Such content is flagged as privacy-sensitive, enabling the Privacy Policy Engine to decide whether visual information may leave the trusted home environment.
- **Audio Analysis Module.** Processes speech content through automatic speech recognition and natural language understanding. Detects personally identifiable information (names, addresses, medical details) and emotional states in conversations. As spoken utterances frequently reveal sensitive personal data, the module marks such segments as privacy-critical so that downstream layers can redact or suppress them prior to transmission.
- **Activity Recognition Module.** Infers activities of daily living (e.g., eating, sleeping, bathing, medication intake) from fused sensor inputs and classifies the current situation to inform contextual privacy decisions. As several of these activities are inherently sensitive (e.g., bathing) or expose health-related behavioural patterns, the module marks such situations as privacy-critical, allowing the Privacy Policy Engine to apply stricter disclosure constraints.
- **Health State Estimation Module.** Aggregates physiological measurements and behavioural indicators to estimate the user's current health status, including anomaly detection for fall events, vital sign deterioration, or prolonged inactivity. Since health information constitutes a special category of personal data under the GDPR [16], the module explicitly labels these outputs as highly privacy-sensitive, ensuring that subsequent layers enforce appropriate safeguards unless a legitimate context, such as a detected emergency, justifies disclosure.

Each module outputs structured semantic annotations rather than forwarding raw data, thereby creating a machine-readable description of what is happening without necessarily retaining identifiable source material. These annotations include sensitivity labels that the Privacy Policy Engine consumes as the basis for its filtering decisions.

3.5 Privacy Policy Engine

The Privacy Policy Engine is responsible for determining whether a given piece of outgoing information is permissible under the currently active privacy configuration. It operates on three levels of policy specification:

Normative Policies encode legal requirements (e.g., GDPR data minimisation or AI Act principles [16,17]) and institutional guidelines that apply universally regardless of user preferences.

Personal Privacy Profiles capture user-specific preferences regarding what information may be shared with whom and under which circumstances. These

profiles are configured during an initial guided setup process and can be re-evaluated and refined over time through simple approval/rejection feedback. Predefined profile templates (e.g., *conservative*, *balanced*, *open*) reduce configuration complexity for users with limited digital literacy.

Contextual Rules dynamically modulate policies based on the current situation, following the principle of contextual integrity [8,38,20]. For example, during a scheduled telemedicine consultation, the system may temporarily permit video transmission of the user's face, while blocking the same information during an unscheduled call from an unknown contact. Similarly, in a detected emergency (e.g., a fall), privacy constraints may be automatically relaxed to enable rapid assistance.

The policy engine employs a rule-based reasoning component combined with a lightweight large language model (LLM) for natural-language policy interpretation and conflict resolution. It consumes the sensitivity labels produced by the Semantic Analysis Layer as a primary input for risk scoring. When multiple rules apply simultaneously, a priority hierarchy and contextual weighting determine the final decision. Each decision is assigned a confidence score; low-confidence cases are either handled conservatively (default to blocking) or escalated to a caregiver for confirmation.

3.6 Data Transformation Layer

Once the Privacy Policy Engine has determined that certain information requires modification before transmission, the Data Transformation Layer applies the appropriate privacy-preserving operation. The following transformation strategies are supported:

- **Blocking.** Complete suppression of the data stream. No information is forwarded to the external service.
- **Redaction.** Selective removal of privacy-sensitive elements while preserving the remainder (e.g., removing spoken names from an audio stream, masking a document visible in a camera frame).
- **Anonymisation.** Replacement of identifiable features with non-identifiable representations (e.g., face blurring, voice modification, replacement of a person's image with an abstract avatar).
- **Aggregation.** Conversion of granular time-series data into summarised indicators (e.g., transmitting "active for 6 hours today" instead of a continuous motion trajectory).
- **Abstraction.** Semantic elevation of information to a less specific level (e.g., reporting "vital signs within normal range" instead of exact heart rate and blood pressure values).

The appropriate transformation is selected based on the policy decision and the minimum information requirement of the receiving service. Where technically feasible, transformations are applied in real time to avoid buffering sensitive raw data.

3.7 Output Gateway and Explainability

The Output Gateway serves as the final control point before information leaves the home environment. It performs three functions:

1. **Transmission Control.** Forwards only approved and transformed data to the designated external recipient via encrypted channels.
2. **Decision Logging.** Maintains a local, privacy-preserving audit log documenting what information was transmitted, blocked, or transformed, along with the corresponding policy justification.
3. **Explanation Generation.** Produces human-readable explanations of filtering decisions, accessible through a simplified user interface. Explanations are formulated in non-technical language (e.g., "Your face was blurred during the video call because your profile does not allow facial images to be shared with this contact.").

The explanation interface is designed for multiple stakeholders: older adults receive simplified summaries, while caregivers and healthcare professionals can access more detailed decision reports when needed.

3.8 Deployment and Hardware Considerations

The entire Personal AI Firewall is designed to operate on local edge computing hardware positioned within the user's home (e.g., a dedicated gateway device or a small-form-factor server). This ensures that raw sensor data never leave the premises, even temporarily.

Current-generation edge AI accelerators (e.g., devices based on NPU or GPU architectures with 10–40 TOPS of inference performance) are capable of executing the required deep learning models for visual, auditory, and behavioural analysis with acceptable latency [56,40]. Model optimisation techniques such as quantisation, pruning, and knowledge distillation are employed [40,62,25] to ensure real-time operation within the power and thermal constraints of a domestic deployment.

Network integration is achieved by positioning the firewall device as the mandatory gateway for all outgoing data traffic from AAL-related systems, analogous to a traditional network firewall but operating at the semantic content level rather than the packet level.

3.9 User Interaction and Onboarding

Recognising that older adults represent the primary user group, the interaction design follows established principles from gerontechnology research [29,3,50]. The initial onboarding process uses guided interviews, preference elicitation through concrete scenario examples ("Would you like your neighbour to see your living room during a video call?"), and progressive disclosure of advanced settings.

Ongoing interaction is kept minimal. The system operates autonomously under normal conditions and only requests user input in ambiguous situations or

when policy conflicts arise. A simplified dashboard, accessible via tablet, television overlay, or voice interface, provides a periodic summary of privacy-relevant events and allows one-touch approval or rejection of specific decisions, which in turn refines the personal privacy profile over time.

4 Proposed Evaluation Methodology

The present work introduces the Personal AI Firewall as a conceptual architecture. A full empirical validation is the subject of ongoing work. In this section we therefore define a structured evaluation methodology that is intended to guide the assessment of a future prototype. The methodology is organised along three complementary dimensions that reflect the socio-technical nature of the system: *technical feasibility* (Section 4.3), *privacy effectiveness* (Section 4.4), and *user acceptance and explainability* (Section 4.5). An illustrative walkthrough (Section 4.6) demonstrates how the layers interact for a concrete scenario.

4.1 Evaluation Dimensions and Research Questions

The evaluation is driven by three guiding research questions:

RQ1 (Feasibility). Can the semantic analysis, policy reasoning, and data transformation operate within the latency, energy, and resource budget of a domestic edge gateway?

RQ2 (Effectiveness). To what extent does the system reduce the disclosure of privacy-sensitive information while preserving the utility required by the intended service?

RQ3 (Acceptance). Do older adults and caregivers understand, trust, and accept the automated privacy decisions and their explanations?

4.2 Datasets

Because the Personal AI Firewall operates on heterogeneous modalities, the evaluation draws on established, publicly available benchmark datasets, selected to cover each analysis module of the Semantic Analysis Layer. Table 2 summarises the intended data sources per modality and evaluation purpose. Datasets requiring a credentialed data-use agreement (e.g., MIMIC) are accessible through the institutional research licence.

The use of standard public benchmarks ensures reproducibility and comparability with related work, while no personal data of the target user group are collected during the technical evaluation phase.

4.3 Dimension 1: Technical Feasibility

The firewall must not degrade the responsiveness of time-critical AAL services (e.g., fall alerting). We propose to deploy the prototype on representative edge hardware (NPU/GPU platforms in the 10–40 TOPS range, cf. Section 3) and to measure:

Table 2: Public benchmark datasets foreseen for the evaluation, grouped by modality and firewall analysis module.

Modality / Module	Dataset	Evaluation purpose
Vision – person/face detection	WIDER FACE [63], LFW [23]	Detection of privacy-critical persons/faces; robustness of anonymisation
Vision – attribute/scene	CelebA [33]	Attribute-level analysis; masking of identifiable features
Activity / ADL	CASAS [13], UCI-HAR [27], Opportunity [14]	Activity recognition; aggregation and abstraction quality
Fall detection	SisFall [55], UP-Fall [36]	Emergency detection under privacy-preserving transformation
Audio / speech	LibriSpeech [43] (PII in transcripts), VoxCeleb [41] (speaker identity)	Detection of spoken PII; speech redaction and speaker anonymisation
Vital signs / health	WESAD [52], PPG-DaLiA [7], MIMIC-IV [26]	Health-state estimation; abstraction of physiological values

- **End-to-end latency.** The additional delay (in ms) introduced by the full pipeline (ingestion → semantic analysis → policy engine → transformation → output), reported per modality and for the combined multimodal case.
- **Resource footprint.** CPU/NPU utilisation, peak memory, and power draw (W) under continuous operation, to verify compliance with the power and thermal constraints of a domestic device.
- **Detection quality of the Semantic Analysis Layer.** Precision, recall, and F1-score for the identification of privacy-sensitive content (recognisable faces on WIDER FACE/LFW, spoken PII on LibriSpeech transcripts, activities on CASAS/UCI-HAR, fall events on SisFall/UP-Fall), reported per detector.

The impact of model-optimisation techniques (quantisation, pruning, knowledge distillation) is assessed by comparing latency and detection quality before and after optimisation, thereby quantifying the accuracy–efficiency trade-off on the edge.

Contemporary edge accelerators (e.g., NVIDIA Jetson Orin-series or comparable NPU platforms) can execute quantised object/face detectors of the YOLO family [47] at interactive frame rates on such hardware [56,40], and an opti-

mised automatic speech-recognition model faster than real time on streaming audio [46,18,45]. Activity and health estimation from wearable and environmental time-series are computationally light by comparison (typically a few milliseconds per inference window). Under an event-triggered scheduling strategy (Section 5.2), the heavy visual and audio models are not run continuously but activated on demand, so that the sustained load is dominated by the lightweight always-on detectors. The principal open question is therefore not the feasibility of any single model, but whether the *combined* pipeline, including the policy engine's language-model reasoning step, can remain within an interactive latency budget (on the order of 300–500 ms for non-safety-critical flows to maintain the user's flow of thought without perceived sluggishness [42,6], and considerably tighter for emergency alerting). Quantifying this combined, concurrent load on representative hardware is a central objective of the proposed evaluation. These figures are indicative estimates drawn from published benchmarks rather than own measurements, and serve only to establish first-order plausibility.

4.4 Dimension 2: Privacy Effectiveness

The core contribution of the firewall is the semantic reduction of disclosed information. We propose two complementary measures:

- **Privacy gain.** The proportion of identifiable elements that are successfully removed or transformed before transmission. This is operationalised through adversarial re-identification attempts on the transformed output, for example, the residual face re-identification rate on LFW after anonymisation, or speaker verification accuracy on VoxCeleb after voice modification. A lower residual identifiability indicates higher privacy gain.
- **Privacy–utility trade-off.** For each transformation strength, we quantify the residual utility for the intended service. For instance, we measure whether a fall can still be reliably detected on SisFall/UP-Fall after body-region abstraction, or whether health status remains assessable after physiological values are abstracted to categorical ranges (WESAD/PPG-DaLiA).

4.5 Dimension 3: User Acceptance and Explainability

Because the target group consists of older adults, technical metrics alone are insufficient. We propose formative user studies with two stakeholder groups: (a) older adults as primary users and (b) caregivers and healthcare professionals as secondary users. Using scenario-based walkthroughs (cf. Section 3), we assess:

1. **Usability**, using the System Usability Scale (SUS) [9] and task-completion measures for the onboarding and approval workflows.
2. **Trust and perceived control**, using a questionnaire capturing users' confidence in the system's decisions and their sense of informational self-determination.
3. **Explanation quality**, evaluated by asking participants to correctly interpret *why* a given decision (e.g., a blurred face during a call) was made. Comprehension accuracy serves as a proxy for the effectiveness of the explanation interface.

4.6 Illustrative Walkthrough

To make the interaction of the layers concrete, we consider a representative scenario: a scheduled telemedicine consultation during which a sensitive document happens to lie on the table within the camera's field of view.

1. The *Data Ingestion Layer* captures the video stream and the scheduling context.
2. The *Semantic Analysis Layer* detects the user's face and, separately, a document region classified as privacy-critical.
3. The *Privacy Policy Engine* resolves the context: the contact is an authorised physician during a scheduled call, so facial transmission is permitted, whereas the document is not necessary for the service and is therefore flagged for removal.
4. The *Data Transformation Layer* preserves the face but masks the document region in real time.
5. The *Output Gateway* forwards the transformed stream, logs the decision, and generates the explanation: *"A document visible on your table was hidden because it is not needed for this medical call."*

5 Discussion

The proposed Personal AI Firewall introduces a new paradigm for privacy protection in Ambient Assisted Living. As a conceptual contribution, however, it raises a number of ethical, technical, and legal questions that must be critically examined before deployment. This section discusses the most salient of these challenges.

5.1 The Break-Glass Dilemma

The most fundamental tension in the proposed system arises from the conflict between its two central objectives: minimising the disclosure of sensitive information on the one hand, and safeguarding the health and safety of the user on the other. A firewall that is configured to be maximally protective may, in the worst case, suppress a safety-critical signal, for example, by classifying the video of a collapsed user as an "exposed body" and therefore blocking its transmission to emergency responders. In a health-critical context, such a false positive is not merely an inconvenience but a potentially life-threatening failure.

To resolve this dilemma, the Contextual Rules of the Privacy Policy Engine (Section 3.5) must implement an explicit *break-glass* mechanism [15,64,34]. Upon detection of a critical event by the Health State Estimation Module (e.g., a fall or a marked deterioration of vital signs), privacy constraints are automatically relaxed to the extent necessary to enable timely assistance. The design of this mechanism must weigh two opposing error types:

- **False positives** (unnecessary disclosure): a non-critical situation is misinterpreted as an emergency, causing sensitive data to be released without genuine need.
- **False negatives** (suppressed alert): a genuine emergency is not recognised as such, and the protective logic wrongly withholds information that is required to help the user.

We argue that in health-critical AAL settings the system should adopt a *fail-safe* default that errs toward the preservation of life: where the confidence of the emergency classifier is uncertain, the system should favour disclosure to a trusted emergency recipient rather than protection. This asymmetry deliberately trades a controlled and auditable privacy loss against the avoidance of catastrophic outcomes. Crucially, every break-glass activation must be recorded in the audit log (Section 3.7) together with its justification, and the user or their caregiver must be notified after the fact, thereby preserving transparency and enabling retrospective review even when prior consent could not be obtained.

5.2 Limitations of Edge-Based Processing

Locating the entire pipeline on edge hardware is essential to the privacy premise of the system, but it imposes significant constraints. Continuously executing multiple deep learning models, for visual, auditory, activity, and health analysis, in parallel is computationally demanding and results in sustained power consumption and heat generation that are difficult to accommodate in an unobtrusive domestic device. Passive cooling, limited physical space, and the expectation of silent, always-on operation further restrict the available thermal and energy budget.

Several mitigations are foreseen. Model-optimisation techniques such as quantisation, pruning, and knowledge distillation (Section 3.8) reduce the computational footprint at the cost of a controlled loss in accuracy, a trade-off that the proposed evaluation (Section 4) is designed to quantify. Furthermore, an *event-triggered* inference strategy can markedly reduce the average load: lightweight always-on detectors continuously monitor the sensor streams, while the more expensive semantic models are activated only when a potentially relevant event is detected. Nevertheless, this strategy introduces its own risk, as the sensitivity of the triggering stage directly bounds the system's ability to detect rare but critical events. Balancing continuous protection against resource efficiency therefore remains an open engineering challenge.

5.3 Reliability of AI-Based Semantic Decisions

The firewall delegates consequential decisions to machine learning components, including a large language model used for policy interpretation and conflict resolution. This raises concerns regarding robustness, bias, and predictability. Detection models may fail on under-represented situations, unusual home layouts, or atypical behaviour, and an LLM-based reasoning component may produce

Fig. 2: Semantic abstraction reduces personal information before transmission, replacing identifiable visual, textual, and acoustic content with less identifiable, service-sufficient representations while preserving utility.

inconsistent or unjustified decisions, particularly in ambiguous cases. Because the target group is especially vulnerable, such failures carry a higher cost than in general-purpose applications.

The concept mitigates these risks through the *privacy-by-default* principle and confidence-based escalation (Section 3.5): low-confidence decisions default to the conservative action or are referred to a caregiver for confirmation. This shifts the residual risk from an uncontrolled disclosure towards an over-restriction that a human can subsequently relax. However, the reliance on human fallback partially reintroduces the very burden the system seeks to remove, and the appropriate division of labour between autonomous operation and human oversight requires empirical study with real users.

5.4 Responsibility, Trust, and Accountability

By actively deciding which information leaves the home, the Personal AI Firewall assumes a role that carries considerable responsibility. Figure 2 illustrates how the same content can be abstracted to varying degrees before it leaves the home (cf. Section 3.6). If a filtering decision contributes to harm, for instance, by delaying assistance, the question of liability arises: is it attributable to the device manufacturer, the operator of the AAL service, the caregiver who configured the profile, or the user who accepted a given privacy template? Existing privacy mechanisms in AAL largely defer to organisational policies, legal regulation such as the GDPR [16], and trust in service providers [53]; a system that autonomously governs data disclosure blurs these established lines of accountability.

We do not claim to resolve this question, but we emphasise two design commitments that are prerequisites for responsible operation. First, *transparency*: the local audit log and the human-readable explanations (Section 3.7) ensure that every decision is traceable and contestable after the fact. Second, *meaningful human oversight*: the system is positioned as an assistant that reduces routine privacy decisions rather than as a fully autonomous authority, retaining caregivers and users in the loop for critical and ambiguous cases. Establishing

the legal and regulatory framework within which such a semantic gateway may operate remains an important direction for interdisciplinary future work.

6 Conclusion and Outlook

This paper introduced the concept of a Personal AI Firewall, a semantic privacy gateway that repositions privacy protection in Ambient Assisted Living from network-level security towards content-aware, AI-driven data governance. In contrast to conventional firewalls and privacy-preserving machine-learning methods, which respectively regulate communication channels and the location of computation, the proposed system continuously evaluates *which* information should leave the home, based on its semantic content and situational context. We presented five guiding design principles, a layered edge architecture comprising data ingestion, semantic analysis, a context-aware policy engine, data transformation, and an explaining output gateway, as well as a structured methodology for its future evaluation along the dimensions of technical feasibility, privacy effectiveness, and user acceptance.

The central contribution of this work is conceptual: it defines a continuous, multimodal semantic gateway tailored to the needs of older adults. Whereas prior work governs either *where* data may flow, network-layer firewalls that filter by traffic destination and treat content inspection as out of scope [35] or *how* it is computed, privacy-preserving learning and edge processing, and while LLM-based policy assistants remain confined to predefined sensor streams within a single cooperative environment [59], the Personal AI Firewall instead continuously evaluates *what* information should leave the home, independent of its source, unifying what were previously isolated, single-modality privacy filters. By combining privacy-by-default, minimal disclosure, and transparent, explainable decisions, the concept aims to reconcile the benefits of digital health monitoring with the right to informational self-determination, without imposing a configuration burden on its users.

As this contribution is conceptual, the immediate next step is the implementation of a prototype and the empirical validation outlined in Section 4, including the measurement of end-to-end latency, detection quality, and the privacy–utility trade-off on representative edge hardware, complemented by formative studies with older adults and caregivers. Looking ahead, personal firewalls are expected to safeguard user privacy against cybercriminals, scams, and social engineering attacks, while preventing vulnerable demographics, such as older adults, from misplaced trust in strangers. Beyond this, we identify two principal directions for future work. First, interoperability: integrating the firewall with emerging smart-home standards such as *Matter* and *Thread* [12,57] would allow it to act as a mandatory semantic gateway across heterogeneous devices without vendor-specific adaptation. Second, reasoning capability: the increasing availability of capable *on-device* Large Language Models opens the possibility of richer contextual policy interpretation and more natural user interaction [62,25], while preserving the strict on-premises processing that underpins the system's privacy

guarantees. Pursuing these directions, together with an interdisciplinary treatment of the legal and ethical questions raised in Section 5, is essential to advance the Personal AI Firewall from a concept towards a deployable building block for trustworthy Ambient Assisted Living.

7 Acknowledgements

We thank the AI tool FhGenie [60] for linguistic support in improving the expression, spelling, and clarity of this work.
This research was funded by the German Federal Ministry for Economic Affairs and Energy (Project KK6143601SK5 GentleCare) and by Fraunhofer Society, Germany.

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L.: Deep Learning with Differential Privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. pp. 308–318. ACM, Vienna Austria (Oct 2016). https://doi.org/10.1145/2976749.2978318, https://dl.acm.org/doi/10.1145/2976749.2978318
2. Acar, A., Aksu, H., Uluagac, A.S., Conti, M.: A survey on homomorphic encryption schemes: Design and challenges. ACM Computing Surveys (CSUR) **51**(5), 1–35 (2018)
3. Aleti, T., Figueiredo, B., Reid, M., Martin, D.M., Sheahan, J., Hjorth, L.: Older adults' digital competency, digital risk perceptions and frequency of everyday digital engagement. Information Technology & People **38**(8), 97–118 (Dec 2025). https://doi.org/10.1108/ITP-05-2024-0624, https://www.emerald.com/itp/article/38/8/97/1267282/Older-adults-digital-competency-digital-risk
4. Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J.M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., Herrera, F.: Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Information Fusion **99**, 101805 (Nov 2023). https://doi.org/10.1016/j.inffus.2023.101805, https://linkinghub.elsevier.com/retrieve/pii/S1566253523001148
5. Alirezaie, M., Renoux, J., Köckemann, U., Kristoffersson, A., Karlsson, L., Blomqvist, E., Tsiftes, N., Voigt, T., Loutfi, A.: An Ontology-based Context-aware System for Smart Homes: E-care@home. Sensors **17**(7), 1586 (Jul 2017). https://doi.org/10.3390/s17071586, https://www.mdpi.com/1424-8220/17/7/1586
6. Attig, C., Rauh, N., Franke, T., Krems, J.F.: System Latency Guidelines Then and Now – Is Zero Latency Really Considered Necessary? In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics: Cognition and Design. pp. 3–14. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-58475-1_1
7. Attila Reiss, I.I.: PPG-DaLiA (2019). https://doi.org/10.24432/C53890, https://archive.ics.uci.edu/dataset/495
8. Bagdasarian, E., Yi, R., Ghalebikesabi, S., Kairouz, P., Gruteser, M., Oh, S., Balle, B., Ramage, D.: AirGapAgent: Protecting Privacy-Conscious Conversational

Agents. In: Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security. pp. 3868–3882. CCS '24, Association for Computing Machinery, New York, NY, USA (Dec 2024). https://doi.org/10.1145/3658644.3690350, https://dl.acm.org/doi/10.1145/3658644.3690350

9. Brooke, J.: SUS: A quick and dirty usability scale. Usability Eval. Ind. **189** (Nov 1995)
10. Caballero, P., Ortiz, G., Medina-Bulo, I.: Systematic literature review of ambient assisted living systems supported by the Internet of Things. Universal Access in the Information Society **23**(4), 1631–1656 (Nov 2024). https://doi.org/10.1007/s10209-023-01022-w, https://doi.org/10.1007/s10209-023-01022-w
11. Cicirelli, G., Marani, R., Petitti, A., Milella, A., D'Orazio, T.: Ambient Assisted Living: A Review of Technologies, Methodologies and Future Perspectives for Healthy Aging of Population. Sensors **21**(10), 3549 (May 2021). https://doi.org/10.3390/s21103549, https://www.mdpi.com/1424-8220/21/10/3549
12. Connectivity Standards Alliance: Matter Specification (2024), https://csa-iot.org/specs/matter/, published: Online Specification
13. Cook, D.J., Crandall, A.S., Thomas, B.L., Krishnan, N.C.: CASAS: A Smart Home in a Box. Computer **46**(7), 62–69 (Jul 2013). https://doi.org/10.1109/MC.2012.328, http://ieeexplore.ieee.org/document/6313586/
14. Daniel Roggen, A.C.: OPPORTUNITY Activity Recognition (2010). https://doi.org/10.24432/C5M027, https://archive.ics.uci.edu/dataset/226
15. Desai, P., Rajendra, K., Raj: The Break-the-Glass (BtG) principle in Access Control (2012)
16. European Parliament and Council of the European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) (Apr 2016), http://data.europa.eu/eli/reg/2016/679/oj, legislative Body: EP, CONSIL
17. European Parliament and Council of the European Union: Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) (Jun 2024), http://data.europa.eu/eli/reg/2024/1689/oj, legislative Body: CONSIL, EP
18. Feng, C., Lin, Y., Zhuo, S., Su, C., Ramakrishnan, R.K., Yuan, Z., Zhang, X.: Edge-ASR: Towards Low-Bit Quantization of Automatic Speech Recognition Models (2025). https://doi.org/10.48550/ARXIV.2507.07877, https://arxiv.org/abs/2507.07877, version Number: 2
19. Fu, J., Hong, Y., Ling, X., Wang, L., Ran, X., Sun, Z., Wang, W.H., Chen, Z., Cao, Y.: Differentially Private Federated Learning: A Systematic Review. ACM Computing Surveys **58**(11), 271:1–271:38 (Apr 2026). https://doi.org/10.1145/3801079, https://dl.acm.org/doi/10.1145/3801079
20. Ghalebikesabi, S., Bagdasarian, E., Yi, R., Yona, I., Shumailov, I., Pappu, A., Shi, C., Weidinger, L., Stanforth, R., Berrada, L., Kohli, P., Huang, P.S., Balle, B.: Operationalizing Contextual Integrity in Privacy-Conscious Assistants (2024). https://doi.org/10.48550/ARXIV.2408.02373, https://arxiv.org/abs/2408.02373, version Number: 2

21. Herriger, C., Merlo, O., Eisingerich, A.B., Arigayota, A.R.: Context-contingent privacy concerns and exploration of the privacy paradox in the age of AI, augmented reality, big data, and the internet of things: systematic review. Journal of Medical Internet Research **27**, e71951 (2025)
22. Hua, H., Shang, Z., Li, X., Shi, P., Yang, C., Wang, L., Zhang, P.: Emotional Speech Anonymization: Preserving Emotion Characteristics in Pseudo-speaker Speech Generation. In: 4th Symposium on Security and Privacy in Speech Communication. pp. 55–60. ISCA (Sep 2024). https://doi.org/10.21437/SPSC.2024-10, https://www.isca-archive.org/spsc_2024/hua24_spsc.html
23. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Tech. Rep. 07-49, University of Massachusetts, Amherst (Oct 2007)
24. Hukkelås, H., Lindseth, F.: DeepPrivacy2: Towards Realistic Full-Body Anonymization. In: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1329–1338 (Jan 2023). https://doi.org/10.1109/WACV56688.2023.00138, https://ieeexplore.ieee.org/document/10030153, iSSN: 2642-9381
25. Jiang, S., Zhou, X., Zhang, M., Xu, C., Liao, G., Chen, J., Cao, J.: Edge large language models: a comprehensive survey. CCF Transactions on Pervasive Computing and Interaction **8**(2), 181–210 (Jun 2026). https://doi.org/10.1007/s42486-025-00227-7, https://link.springer.com/10.1007/s42486-025-00227-7
26. Johnson, A.E.W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T.J., Hao, S., Moody, B., Gow, B., Lehman, L.w.H., Celi, L.A., Mark, R.G.: MIMIC-IV, a freely accessible electronic health record dataset. Scientific Data **10**(1), 1 (Jan 2023). https://doi.org/10.1038/s41597-022-01899-x, https://www.nature.com/articles/s41597-022-01899-x
27. Jorge Reyes-Ortiz, D.A.: Human Activity Recognition Using Smartphones (2013). https://doi.org/10.24432/C54S4K, https://archive.ics.uci.edu/dataset/240
28. Kairouz, P., McMahan, H.B.: Advances and Open Problems in Federated Learning. Foundations and Trends® in Machine Learning **14**(1-2), 1–210 (Jun 2021). https://doi.org/10.1561/2200000083, https://www.emerald.com/ftmal/article/14/1-2/1/1332154/Advances-and-Open-Problems-in-Federated-Learning
29. Knowles, B., Hanson, V.L.: The wisdom of older technology (non)users. Communications of the ACM **61**(3), 72–77 (Feb 2018). https://doi.org/10.1145/3179995, https://dl.acm.org/doi/10.1145/3179995
30. Kung, H.W., Varanka, T., Saha, S., Sim, T., Sebe, N.: Face Anonymization Made Simple. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1040–1050 (Feb 2025). https://doi.org/10.1109/WACV61041.2025.00110, https://ieeexplore.ieee.org/document/10943380, iSSN: 2642-9381
31. Li, Y., Mandalari, A.M., Straw, I.: Who Let the Smart Toaster Hack the House? An Investigation into the Security Vulnerabilities of Consumer IoT Devices. In: Proceedings of the 2023 International Conference on embedded Wireless Systems and Networks. pp. 415–420. EWSN '23, International Conference on Embedded Wireless Systems and Networks (EWSN), Rende, Italy (Dec 2023)
32. Lindell, Y.: Secure multiparty computation (MPC). Communications of the ACM **63**(1), 86–96 (2020)

33. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep Learning Face Attributes in the Wild. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 3730–3738. IEEE, Santiago, Chile (Dec 2015). https://doi.org/10.1109/ICCV.2015.425, http://ieeexplore.ieee.org/document/7410782/
34. Loos, M.: Break-Glass Access Control Systems in Medical Devices (Jan 2021). https://doi.org/10.13140/RG.2.2.10317.51684, researchGate Preprint
35. Mandalari, A.M., Dubois, D.J., Kolcun, R., Paracha, M.T., Haddadi, H., Choffnes, D.: Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic (May 2021). https://doi.org/10.48550/arXiv.2105.05162, http://arxiv.org/abs/2105.05162, arXiv:2105.05162 [cs.NI]
36. Martínez-Villaseñor, L., Ponce, H., Brieva, J., Moya-Albor, E., Núñez-Martínez, J., Peñafort-Asturiano, C.: UP-Fall Detection Dataset: A Multimodal Approach. Sensors **19**(9), 1988 (Apr 2019). https://doi.org/10.3390/s19091988, https://www.mdpi.com/1424-8220/19/9/1988
37. Mendez, J., Bierzynski, K., Cuéllar, M.P., Morales, D.P.: Edge Intelligence: Concepts, Architectures, Applications, and Future Directions. ACM Transactions on Embedded Computing Systems (TECS) **21**(5), 1–36 (2022). https://doi.org/10.1145/3486674
38. Mireshghallah, N., Kim, H., Zhou, X., Tsvetkov, Y., Sap, M., Shokri, R., Choi, Y.: Can LLMs Keep a Secret? Testing Privacy Implications of Language Models via Contextual Integrity Theory. International Conference on Learning Representations **2024**, 1892–1915 (May 2024), https://proceedings.iclr.cc/paper_files/paper/2024/hash/08305d8b2ddab98932c163ea73df065f-Abstract-Conference.html
39. Morel, V., Iwaya, L., Fischer-Hübner, S.: AI-driven Personalized Privacy Assistants: a Systematic Literature Review. IEEE Access **13**, 160982–161002 (2025). https://doi.org/10.1109/ACCESS.2025.3609188, http://arxiv.org/abs/2502.07693, arXiv:2502.07693 [cs.CY]
40. Murshed, M.G.S., Murphy, C., Hou, D., Khan, N., Ananthanarayanan, G., Hussain, F.: Machine Learning at the Network Edge: A Survey. ACM Computing Surveys **54**(8), 1–37 (Nov 2022). https://doi.org/10.1145/3469029, https://dl.acm.org/doi/10.1145/3469029
41. Nagrani, A., Chung, J.S., Zisserman, A.: VoxCeleb: A Large-Scale Speaker Identification Dataset. In: Interspeech 2017. pp. 2616–2620. ISCA (Aug 2017). https://doi.org/10.21437/Interspeech.2017-950, https://www.isca-archive.org/interspeech_2017/nagrani17_interspeech.html
42. Nielsen, J.: Usability engineering. Academic Press, Boston (1993)
43. Panayotov, V., Chen, G., Povey, D., Khudanpur, S.: Librispeech: An ASR corpus based on public domain audio books. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 5206–5210. IEEE, South Brisbane, Queensland, Australia (Apr 2015). https://doi.org/10.1109/ICASSP.2015.7178964, http://ieeexplore.ieee.org/document/7178964/
44. Pathmabandu, C., Grundy, J., Chhetri, M.B., Baig, Z.: Privacy for IoT: Informed consent management in Smart Buildings. Future Generation Computer Systems **145**, 367–383 (Aug 2023). https://doi.org/10.1016/j.future.2023.03.045, https://www.sciencedirect.com/science/article/pii/S0167739X23001322
45. Qin, R., Liu, D., Xu, G., Nassereldine, A., Yan, Z., Xu, C., Hu, Y., Hu, X.S., Xiong, J., Shi, Y.: Tiny-Align: Bridging Automatic Speech Recognition and Large Language Model on Edge. In: 2025 IEEE/ACM International Conference On Computer Aided Design (ICCAD). pp. 1–9 (Oct 2025). https://

doi.org/10.1109/ICCAD66269.2025.11240668, https://ieeexplore.ieee.org/abstract/document/11240668, iSSN: 1558-2434
46. Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I.: Robust speech recognition via large-scale weak supervision. In: Proceedings of the 40th International Conference on Machine Learning. ICML'23, vol. 202, pp. 28492–28518. JMLR.org, Honolulu, Hawaii, USA (Jul 2023)
47. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 779–788 (Jun 2016). https://doi.org/10.1109/CVPR.2016.91, https://ieeexplore.ieee.org/document/7780460, iSSN: 1063-6919
48. Ren, J., Dubois, D.J., Choffnes, D., Mandalari, A.M., Kolcun, R., Haddadi, H.: Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach. In: Proceedings of the Internet Measurement Conference. pp. 267–279. ACM, Amsterdam Netherlands (Oct 2019). https://doi.org/10.1145/3355369.3355577, https://dl.acm.org/doi/10.1145/3355369.3355577
49. Saeed, W., Omlin, C.: Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. Knowledge-Based Systems **263**, 110273 (Jan 2023). https://doi.org/10.1016/j.knosys.2023.110273
50. Saka, S., Das, S.: "Watch My Health, Not My Data": Understanding Perceptions, Barriers, Emotional Impact, & Coping Strategies Pertaining to IoT Privacy and Security in Health Monitoring for Older Adults. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. pp. 1–18. CHI '25, Association for Computing Machinery, New York, NY, USA (Apr 2025). https://doi.org/10.1145/3706598.3714019, https://dl.acm.org/doi/10.1145/3706598.3714019
51. Saka, S., Das, S.: SoK: Reviewing Two Decades of Security, Privacy, Accessibility, and Usability Studies on Internet of Things for Older Adults. In: Proceedings of the ACM Asia Conference on Computer and Communications Security. pp. 98–115. ACM, Bangalore India (Jun 2026). https://doi.org/10.1145/3779208.3785270, https://dl.acm.org/doi/10.1145/3779208.3785270
52. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., Van Laerhoven, K.: Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction. pp. 400–408. ACM, Boulder CO USA (Oct 2018). https://doi.org/10.1145/3242969.3242985, https://dl.acm.org/doi/10.1145/3242969.3242985
53. Stach, C., Gritti, C., Bräcker, J., Behringer, M., Mitschang, B.: Protecting Sensitive Data in the Information Age: State of the Art and Future Prospects. Future Internet **14**(11), 302 (Nov 2022). https://doi.org/10.3390/fi14110302, https://www.mdpi.com/1999-5903/14/11/302
54. Stenger, R., Busse, S., Sander, J., Eisenbarth, T., Fudickar, S.: Evaluating the Impact of Face Anonymization Methods on Computer Vision Tasks: A Trade-Off Between Privacy and Utility. IEEE Access **13**, 11070–11079 (2025). https://doi.org/10.1109/ACCESS.2024.3519441, https://ieeexplore.ieee.org/document/10804775
55. Sucerquia, A., López, J., Vargas-Bonilla, J.: SisFall: A Fall and Movement Dataset. Sensors **17**(1), 198 (Jan 2017). https://doi.org/10.3390/s17010198, https://www.mdpi.com/1424-8220/17/1/198
56. Süzen, A.A., Duman, B., Şen, B.: Benchmark Analysis of Jetson TX2, Jetson Nano and Raspberry PI using Deep-CNN. In: 2020 International Congress on

Human-Computer Interaction, Optimization and Robotic Applications (HORA). pp. 1–5 (Jun 2020). https://doi.org/10.1109/HORA49412.2020.9152915, https://ieeexplore.ieee.org/document/9152915
57. Thread Group: Thread Protocol Specification (2024), https://www.threadgroup.org/, published: Online Specification
58. Tomashenko, N., Miao, X., Champion, P., Meyer, S., Panariello, M., Wang, X., Evans, N., Vincent, E., Yamagishi, J., Todisco, M.: The third VoicePrivacy challenge: Preserving emotional expressiveness and linguistic content in voice anonymization. Computer Speech & Language **100**, 101988 (Oct 2026). https://doi.org/10.1016/j.csl.2026.101988, https://www.sciencedirect.com/science/article/pii/S0885230826000513
59. Wang, B., Garcia, L.A., Srivastava, M.: PrivacyOracle: Configuring Sensor Privacy Firewalls with Large Language Models in Smart Built Environments. In: 2024 IEEE Security and Privacy Workshops (SPW). pp. 239–245 (May 2024). https://doi.org/10.1109/SPW63631.2024.00028, https://ieeexplore.ieee.org/document/10579527, iSSN: 2770-8411
60. Weber, I., Linka, H., Mertens, D., Muryshkin, T., Opgenoorth, H., Langer, S.: Fhgenie: a custom, confidentiality-preserving chat ai for corporate and scientific use. In: 2024 IEEE 21st International Conference on Software Architecture Companion (ICSA-C). pp. 26–31. IEEE (2024). https://doi.org/10.1109/icsa-c63560.2024.00011
61. Xin, W., Jiaqian, L., Xueshuang, D., Haoji, Z., Lianshan, S.: A Survey of Differential Privacy Techniques for Federated Learning. IEEE Access **13**, 6539–6555 (2025). https://doi.org/10.1109/ACCESS.2024.3523909, https://ieeexplore.ieee.org/document/10818489
62. Xu, J., Li, Z., Chen, W., Wang, Q., Gao, X., Cai, Q., Ling, Z.: On-Device Language Models: A Comprehensive Review (2024). https://doi.org/10.48550/ARXIV.2409.00088, https://arxiv.org/abs/2409.00088, version Number: 2
63. Yang, S., Luo, P., Loy, C.C., Tang, X.: WIDER FACE: A Face Detection Benchmark. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5525–5533. IEEE, Las Vegas, NV, USA (Jun 2016). https://doi.org/10.1109/CVPR.2016.596, https://ieeexplore.ieee.org/document/7780965/
64. Yang, Y., Liu, X., Deng, R.H.: Lightweight Break-Glass Access Control System for Healthcare Internet-of-Things. IEEE Transactions on Industrial Informatics **14**(8), 3610–3617 (Aug 2018). https://doi.org/10.1109/TII.2017.2751640, https://ieeexplore.ieee.org/abstract/document/8036213
65. Zhang, Z., Cilloni, T., Walter, C., Fleming, C.: Multi-Scale, Class-Generic, Privacy-Preserving Video. Electronics **10**(10), 1172 (Jan 2021). https://doi.org/10.3390/electronics10101172, https://www.mdpi.com/2079-9292/10/10/1172
66. Zieni, B., Ritchie, M.A., Mandalari, A.M., Boem, F.: An Interdisciplinary Overview on Ambient Assisted Living Systems for Health Monitoring at Home: Trade-Offs and Challenges. Sensors **25**(3), 853 (Jan 2025). https://doi.org/10.3390/s25030853, https://www.mdpi.com/1424-8220/25/3/853