PRIVACY-PRESERVING AI VALIDATION PROTOCOL FOR SENSITIVE DATA

PROVISIONAL PATENT APPLICATION

Inventor: Kinan Lemberg

Address: 270 Bolton Rd, Koah, 4881, Australia

Filing Date: June 3, 2025

FIELD OF THE INVENTION

This invention relates to privacy-preserving validation systems for artificial intelligence applications, and more specifically to protocols that validate AI responses containing sensitive data without storing, transmitting, or exposing the underlying private information through zero-knowledge techniques and homomorphic validation.

BACKGROUND OF THE INVENTION

Current AI validation systems require full access to response content, creating significant privacy risks when handling sensitive medical, financial, or personal data. Healthcare and financial institutions cannot use existing validation systems due to regulatory requirements prohibiting data exposure. This creates a critical gap where the most sensitive AI applications lack proper validation.

Current limitations include:

- No validation methods that preserve data privacy
- Lack of zero-knowledge proof systems for AI validation
- Absence of homomorphic validation techniques
- No compliance with GDPR, HIPAA, and financial privacy regulations
- Limited ephemeral processing capabilities for sensitive content

There exists a need for a comprehensive privacy-preserving validation protocol that can validate AI responses containing sensitive data without accessing, storing, or transmitting the private information.

SUMMARY OF THE INVENTION

The present invention provides a privacy-preserving AI validation protocol that validates responses containing sensitive data through zero-knowledge proofs, homomorphic encryption, and ephemeral processing techniques while maintaining full regulatory compliance.

The invention comprises:

1. Zero-Knowledge Validation Engine - Validation of AI responses without revealing content through cryptographic zero-knowledge proofs.

2. Homomorphic Processing System - Mathematical validation operations on encrypted data without decryption requirements.

3. Ephemeral Computation Framework - Temporary processing environment with guaranteed data destruction after validation.

4. Privacy-Compliant Audit System - Audit trail generation that proves validation occurred without storing sensitive content.

5. Regulatory Compliance Module - Automated compliance verification for GDPR, HIPAA, PCI-DSS, and other privacy regulations.

The system provides significant advantages by enabling validation of sensitive AI responses while maintaining complete data privacy and regulatory compliance.

DETAILED DESCRIPTION OF THE INVENTION

System Architecture

The Privacy-Preserving AI Validation Protocol operates as a cryptographic layer that validates AI responses containing sensitive data without exposing the underlying private information.

1. Zero-Knowledge Validation Engine

The Zero-Knowledge Engine implements cryptographic protocols allowing validation proof without content revelation.

Zero-Knowledge Proof Construction:

```
ZK_Validation_Protocol = {
    commitment_phase: {
        response_commitment: Commit(AI_Response, Random_Nonce),
        validation_criteria: Public_Validation_Requirements,
        proof_structure: Interactive_or_Non_Interactive_ZKP,
        binding_property: Computationally_Binding_Commitment,
        hiding_property: Perfectly_Hiding_Commitment
    },
    proof_generation: {
        statement: "Response R satisfies validation criteria V",
        witness: Private_Response_Content,
        proof: ZK_SNARK_or_ZK_STARK_Proof,
        soundness: Negligible_False_Positive_Rate,
        completeness: Always_Accept_Valid_Responses
    },
    verification_phase: {
        public_inputs: Validation_Criteria_and_Commitment,
        proof_verification: Verify(Proof, Public_Inputs),
        result: Accept_or_Reject_Without_Learning_Content,
        verification_time: Constant_Time_Verification
    }
}
```

Validation Criteria Encoding:

```
```

```
Private_Validation_Criteria = {
    accuracy_validation: {
        fact_checking: Merkle_Tree_of_Valid_Facts,
        consistency_rules: Encrypted_Logic_Rules,
        completeness_requirements: Zero_Knowledge_Set_Membership,
        relevance_criteria: Similarity_Threshold_Commitments
    },
    safety_validation: {
        prohibited_content: Bloom_Filter_of_Unsafe_Patterns,
        risk_thresholds: Homomorphic_Risk_Scores,
        compliance_rules: Encrypted_Regulatory_Requirements,
        ethical_guidelines: Private_Set_Intersection_Rules
    }
}

Circuit_Construction = {
    validation_circuit: Boolean_Circuit_for_Validation_Logic,
    arithmetic_circuit: Field_Operations_for_Computations,
    constraint_system: R1CS_or_Plonk_Constraints,
    optimization: Circuit_Size_Minimization,
    trusted_setup: Universal_or_Transparent_Setup
}
```

Interactive vs Non-Interactive Proofs:

```
Interactive_ZKP = {
    prover_verifier_rounds: 3_to_5_Challenge_Response_Rounds,
    fiat_shamir_heuristic: Convert_to_Non_Interactive,
    communication_complexity: O(log n) bits,
    computational_complexity: Polynomial_Time,
    security_parameter: 128_bit_Security_Level
}

Non_Interactive_ZKP = {
    proof_system: Groth16_or_PLONK_or_STARK,
    proof_size: {
        groth16: ~200_bytes,
        plonk: ~400_bytes,
        stark: ~45KB_to_200KB
    },
    verification_time: {
        groth16: ~10ms,
        plonk: ~15ms,
        stark: ~50ms
    },
    trusted_setup_requirement: {
        groth16: Required_Per_Circuit,
        plonk: Universal_Setup,
        stark: No_Trusted_Setup
    }
}
```

2. Homomorphic Processing System

The Homomorphic System enables mathematical validation operations on encrypted data without decryption.

Homomorphic Encryption Schemes:

```
Encryption_Scheme_Selection = {
  fully_homomorphic: {
    scheme: CKKS_or_BGV_or_BFV,
    operations: Arbitrary_Computations,
    performance: High_Computational_Overhead,
    use_case: Complex_Validation_Logic
  },
  partially_homomorphic: {
    scheme: Paillier_or_ElGamal,
    operations: Addition_or_Multiplication_Only,
    performance: Efficient_for_Limited_Ops,
    use_case: Simple_Scoring_and_Thresholds
  },
  somewhat_homomorphic: {
    scheme: BGN_or_Lattice_Based,
    operations: Limited_Depth_Circuits,
    performance: Balanced_Efficiency,
    use_case: Moderate_Complexity_Validation
  }
}

Homomorphic_Validation_Operations = {
  encrypted_scoring: {
    input: Enc(Response_Features),
    computation: Score = Σ(Weight_i × Enc(Feature_i)),
    output: Enc(Validation_Score),
    decryption: Only_by_Authorized_Party
  },
  threshold_comparison: {
    encrypted_score: Enc(Score),
    encrypted_threshold: Enc(Threshold),
    comparison: Enc(Score > Threshold),
        result: Encrypted_Boolean_Result
    },
    pattern_matching: {
        encrypted_text: Character_Wise_Encryption,
        encrypted_patterns: Pattern_Library,
        matching: Homomorphic_String_Comparison,
        efficiency: Optimized_Batching_Techniques
    }
}
```

Noise Management and Bootstrapping:

```
Noise_Growth_Management = {
    initial_noise: Small_Random_Noise_for_Security,
    noise_growth_rate: Multiplicative_per_Operation,
    noise_threshold: Maximum_Before_Decryption_Fails,
    noise_estimation: Track_Noise_Budget_Dynamically
}

Bootstrapping_Strategy = {
    trigger_condition: Noise_Approaching_Threshold,
    bootstrapping_operation: Homomorphic_Decryption_of_Ciphertext,
    performance_impact: 1000x_Slower_Than_Regular_Op,
    optimization: {
        batching: Process_Multiple_Ciphertexts,
        parallelization: Distribute_Across_Cores,
        precomputation: Cache_Bootstrapping_Keys,
        selective: Only_Bootstrap_When_Necessary
    }
}
```

Privacy-Preserving Computations:

```
Secure_Multi_Party_Computation = {
    parties: [User, Validator, Auditor],
    secret_sharing: Shamir_or_Additive_Shares,
    computation_protocol: GMW_or_BGW_or_BMR,
    communication_rounds: O(depth) for circuits,
    security_model: Semi_Honest_or_Malicious
}

Private_Set_Intersection = {
    user_set: Sensitive_Terms_in_Response,
    validator_set: Prohibited_or_Required_Terms,
    protocol: OT_Based_or_Bloom_Filter_PSI,
    output: Cardinality_Only_or_Intersection_Size,
    leakage: Zero_Beyond_Intersection_Size
}
```

3. Ephemeral Computation Framework

The Ephemeral Framework ensures all sensitive data is processed temporarily with guaranteed destruction.

Secure Ephemeral Environment:

```
Trusted_Execution_Environment = {
    hardware_security: Intel_SGX_or_AMD_SEV_or_ARM_TrustZone,
    enclave_properties: {
        isolation: Hardware_Enforced_Memory_Protection,
        attestation: Remote_Attestation_Capability,
        sealing: Secure_Key_Storage,
        memory_encryption: Runtime_Memory_Encryption
    },
    validation_enclave: {
        max_lifetime: 5_minutes_per_validation,
        memory_limit: 4GB_secure_memory,
        no_persistent_storage: RAM_Only_Operations,
        secure_channels: TLS_for_External_Communication
    }
}

Memory_Sanitization = {
    allocation: Secure_Random_Initialization,
    usage_tracking: Monitor_All_Memory_Access,
    deallocation: Cryptographic_Wiping_Pattern,
    verification: {
        overwrite_passes: 3_Pass_DoD_5220_Standard,
        pattern_sequence: [0x00, 0xFF, Random],
        memory_locking: Prevent_Swap_to_Disk,
        verification_read: Confirm_Sanitization_Success
    }
}
```

Ephemeral Key Management:

```
Session_Key_Generation = {
    key_derivation: HKDF_from_Hardware_RNG,
    key_lifetime: Valid_for_Single_Validation,
    key_storage: Enclave_Protected_Memory_Only,
    key_destruction: Immediate_Post_Validation
}

Forward_Secrecy = {
    key_agreement: Ephemeral_ECDH,
    perfect_forward_secrecy: New_Keys_Per_Session,
    key_rotation: Automatic_Mid_Session_if_Needed,
    compromise_resilience: Past_Sessions_Remain_Secure
}

Validation_State_Management = {
    state_initialization: Clean_State_Per_Validation,
    state_isolation: No_State_Sharing_Between_Validations,
    state_destruction: Complete_State_Wipe_Post_Validation,
    audit_preservation: Only_Non_Sensitive_Metadata_Retained
}
```

4. Privacy-Compliant Audit System

The Audit System generates compliance proofs without storing sensitive content.

Privacy-Preserving Audit Records:

```
Audit_Record_Structure = {
    validation_proof: {
        timestamp: Validation_Occurrence_Time,
        proof_hash: Hash(ZK_Proof),
        result: Validation_Pass_or_Fail,
        confidence: Encrypted_Confidence_Score
    },
    metadata_only: {
        request_id: Anonymous_Request_Identifier,
        validation_type: General_Validation_Category,
        processing_time: Duration_in_Milliseconds,
        resource_usage: CPU_Memory_Metrics
    },
    compliance_attestation: {
        regulation: GDPR_HIPAA_PCI_Compliance_Flags,
        data_handling: Privacy_Preservation_Certification,
        retention_policy: Automatic_Deletion_Schedule,
        access_control: Role_Based_Access_List
    }
}

Differential_Privacy_Audit = {
    noise_addition: Laplace_or_Gaussian_Noise,
    privacy_budget: Epsilon_Delta_Parameters,
    query_sensitivity: Maximum_Change_per_Record,
    composition: Sequential_or_Parallel_Composition
}
```

Cryptographic Audit Proofs:

```
Merkle_Tree_Audit = {
    leaf_nodes: Hash(Each_Validation_Event),
    tree_construction: Binary_Merkle_Tree,
    root_commitment: Published_Merkle_Root,
    inclusion_proof: Path_from_Leaf_to_Root,
    non_inclusion_proof: Sorted_Tree_with_Proofs
}

Blockchain_Audit_Trail = {
    private_blockchain: Permissioned_Ledger,
    block_content: {
      header: Block_Metadata_and_Previous_Hash,
      transactions: Validation_Proof_Hashes_Only,
      merkle_root: Root_of_Transaction_Tree,
      signature: Validator_Digital_Signature
    },
    consensus: Practical_Byzantine_Fault_Tolerance,
    privacy_features: {
      confidential_transactions: Pedersen_Commitments,
      range_proofs: Bulletproofs_for_Validity,
      ring_signatures: Hide_Validator_Identity,
      stealth_addresses: One_Time_Audit_Addresses
    }
}
```

5. Regulatory Compliance Module

The Compliance Module ensures automatic adherence to privacy regulations across jurisdictions.

Multi-Regulation Compliance Framework:

```
GDPR_Compliance = {
  data_minimization: Process_Minimum_Required_Data,
  purpose_limitation: Validate_Only_No_Other_Use,
  storage_limitation: Immediate_Deletion_Post_Validation,
  rights_implementation: {
        right_to_erasure: No_Data_to_Erase,
        right_to_access: Provide_Audit_Proof_Only,
        right_to_portability: Export_Non_Sensitive_Records,
        right_to_object: Opt_Out_Capability
    },
    privacy_by_design: {
        default_privacy: Maximum_Privacy_Settings,
        data_protection_impact: Automated_DPIA,
        breach_notification: Impossible_Due_to_Encryption,
        processor_agreements: Cryptographic_Guarantees
    }
}

HIPAA_Compliance = {
    phi_handling: Never_Store_PHI,
    minimum_necessary: Process_Only_Required_Elements,
    access_controls: Role_Based_Cryptographic_Access,
    audit_controls: Comprehensive_Access_Logging,
    transmission_security: End_to_End_Encryption,
    administrative_safeguards: {
        workforce_training: Automated_Compliance_Training,
        access_management: Cryptographic_Key_Management,
        security_officer: Designated_Compliance_Role,
        risk_assessment: Continuous_Risk_Monitoring
    }
}

Financial_Privacy_Compliance = {
    pci_dss: {
        cardholder_data: Never_Stored_or_Transmitted,
        network_security: Encrypted_Channels_Only,
        access_control: Multi_Factor_Authentication,
        monitoring: Real_Time_Security_Monitoring
    },
    sox_compliance: {
        internal_controls: Cryptographic_Control_Proofs,
        audit_trail: Immutable_Validation_Records,
        change_management: Versioned_Validation_Logic,
        segregation_of_duties: Cryptographic_Role_Separation
    }
}
```

Cross-Border Data Protection:

```
Jurisdiction_Detection = {
    user_location: IP_Geolocation_or_Declaration,
    data_residency: Process_in_User_Jurisdiction,
    applicable_laws: Dynamic_Law_Database,
    compliance_selection: Most_Restrictive_Requirements
}

Data_Localization = {
    processing_location: Edge_Computing_in_User_Region,
    key_management: Regional_Key_Management_Services,
    audit_storage: Jurisdiction_Specific_Storage,
    cross_border_transfer: Prohibited_by_Design
}
```

System Integration and Performance

Privacy Performance Requirements:

- ZK Proof Generation: < 100ms for typical validation
- Homomorphic Operations: < 500ms for basic validation
- Ephemeral Processing: < 2s total lifetime
- Audit Record Creation: < 50ms without sensitive data

Privacy Guarantees:

- Information Leakage: Zero beyond validation result
- Storage Requirements: No persistent sensitive data
- Transmission Security: End-to-end encryption
- Computational Privacy: Secure multi-party computation

ADVANTAGES OVER PRIOR ART

The present invention provides significant advantages over existing validation approaches:

1. Zero-Knowledge Validation: Unlike traditional systems, validates without accessing content through cryptographic proofs.

2. Homomorphic Processing: Performs validation on encrypted data rather than requiring plaintext access.

3. Ephemeral Architecture: Guarantees data destruction rather than relying on deletion policies.

4. Privacy-Compliant Auditing: Creates audit trails without sensitive data rather than redacting after storage.

5. Regulatory Compliance: Built-in compliance for multiple regulations rather than retrofitted privacy measures.

6. Healthcare and Financial Enablement: Allows validation in highly regulated industries previously unable to use AI validation.

CLAIMS

Claim 1: A privacy-preserving AI validation protocol comprising:

- a zero-knowledge validation engine proving validation without revealing content;
- a homomorphic processing system performing validation on encrypted data;
- an ephemeral computation framework with guaranteed data destruction;
- a privacy-compliant audit system generating proofs without sensitive content; and
- a regulatory compliance module ensuring automatic multi-regulation adherence.

Claim 2: The protocol of claim 1, wherein the zero-knowledge engine implements zk-SNARKs or zk-STARKs for non-interactive validation proofs.

Claim 3: The protocol of claim 1, wherein the homomorphic system uses CKKS or BGV schemes for encrypted numerical computations.

Claim 4: The protocol of claim 1, wherein the ephemeral framework utilizes trusted execution environments with hardware-enforced isolation.

Claim 5: The protocol of claim 1, wherein the audit system generates merkle tree proofs and blockchain records without sensitive data.

Claim 6: The protocol of claim 1, wherein the compliance module automatically applies GDPR, HIPAA, and PCI-DSS requirements.

Claim 7: A method for privacy-preserving AI validation comprising:

- receiving encrypted AI response requiring validation;
- generating zero-knowledge proof of validation criteria satisfaction;
- performing homomorphic operations for validation scoring;
- processing in ephemeral environment with timed destruction;
- creating privacy-compliant audit record; and
- ensuring regulatory compliance throughout.

Claim 8: The method of claim 7, further comprising using secure multi-party computation for distributed validation without data sharing.

Claim 9: The method of claim 7, wherein ephemeral processing includes cryptographic memory wiping and forward secrecy guarantees.

Claim 10: The method of claim 7, wherein audit records use differential privacy and cryptographic commitments for privacy preservation.

ABSTRACT

A privacy-preserving AI validation protocol enables validation of sensitive AI responses without accessing, storing, or transmitting private data. The system comprises: (1) zero-knowledge validation engine using cryptographic proofs, (2) homomorphic processing for encrypted data validation, (3) ephemeral computation with guaranteed destruction, (4) privacy-compliant audit system without sensitive data storage, and (5) automatic regulatory compliance for GDPR, HIPAA, and financial privacy. The protocol enables AI validation in healthcare, finance, and other regulated industries while maintaining complete data privacy, providing advantages through zero-knowledge proofs, homomorphic encryption, ephemeral processing, and built-in compliance.

END OF PATENT SPECIFICATION