Journal of Theoretical and Applied Information Technology
15st June 2026. Vol.104. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DIGITAL TWIN TRAINING AND LEGAL GOVERNANCE FOR HUMANOID POLICE ROBOTS IN PATROL AND ARREST-SUPPORT

SEUNGKOOK ROH
Department of Data Science, Korean National Police University, Republic of Korea
E-mail: skroh@police.ac.kr

ABSTRACT

Humanoid police robots are emerging as Physical AI platforms because they can operate in human-centered infrastructure such as stairs, doors, corridors, vehicles, and control panels. This study is important because police robots do not merely perform technical tasks; they may mediate state authority, personal data processing, scene preservation, and possible restrictions on bodily liberty. The goal of this paper is to design a digital twin-based training and governance framework for humanoid police robots in patrol and arrest-support operations in South Korea. The central hypothesis is that patrol functions can be trained toward relatively high autonomy only if the digital twin encodes legal permission states, privacy requirements, evidence preservation, and human-command boundaries together with locomotion and perception; conversely, arrest judgment, bodily restraint, and hazardous device activation should not be automated under the current Korean legal framework. The study uses an interdisciplinary design-science method combining a review of digital twin and humanoid learning research, doctrinal analysis of Korean constitutional, criminal procedure, police, privacy, AI, and compensation law, and synthesis of a deployment-oriented architecture. The main findings are a five-layer digital twin, an autonomy-permission matrix, a benchmark-and-phase-gate validation matrix, and a human-in-command operating procedure. Practically, the framework helps police agencies, vendors, and regulators convert broad safety and legality requirements into testable design controls before public pilots. The paper concludes that the most defensible near-term model is not an autonomous arrest robot but a digitally trained police-support humanoid with law-in-the-loop constraints, privacy-by-design controls, and audit-ready evidence logs.

Keywords: *Physical AI, Digital Twin, Humanoid Police Robot, Sim-to-Real Transfer, AI Governance*

1. INTRODUCTION

1.1. Background and Significance

Physical AI, often discussed as embodied artificial intelligence, refers to AI systems that perceive the physical world, reason about it, and act through bodies, sensors, and actuators. Humanoid robots are especially relevant because they can operate in spaces built for humans without requiring major redesign of stairs, doors, corridors, control panels, and vehicles [1]-[5]. In public safety, this ability creates an immediate connection with police work, where officers must move through complex urban environments, monitor anomalies, interact with citizens, preserve scenes, and respond to rapidly changing risk conditions.

Police patrol and arrest-support operations, however, are not ordinary service-robot tasks. They require robust locomotion, social navigation, uncertainty-aware perception, communication under stress, evidence handling, and strict compliance with limits on coercive power. Training a humanoid robot directly in streets, subway stations, airports, campuses, or event venues through repeated trial and error would be unsafe, costly, and socially unacceptable. A digital twin provides a practical route because it couples a physical environment with a data-linked simulation in which robot behavior can be repeated, perturbed, measured, and audited at scale [1]-[10].

The significance of the study lies in the dual character of police robots. A humanoid police robot is simultaneously a learning machine, a mobile sensor, a public-administration instrument, and a potential police equipment platform. A robot that walks efficiently but records people without proper notice,

misclassifies citizens, or initiates force without human authorization would remain unacceptable. Therefore, technical performance and legal validity must be designed together rather than checked separately after deployment.

1.2. Research Gap

The literature reveals a three-part design gap. Technical studies optimize locomotion, manipulation, imitation learning, reinforcement learning, and sim-to-real transfer, but seldom encode public-law limits, evidence preservation, or procedural approval states into the training loop [1]-[10]. Legal scholarship evaluates surveillance, arrest authority, police force, accountability, and public legitimacy, but usually without specifying a trainable robot architecture or measurable deployment criteria [12]-[18]. Pilot-oriented discussions of police robotics identify use cases, yet they rarely provide reproducible baselines, scenario families, phase-gate thresholds, or certification criteria [11], [16]-[18]. Practical deployment therefore requires an integrated architecture that treats robot learning, legal governance, and evidence integrity as one socio-technical system.

1.3. Goal, Hypothesis, and Research Questions

The goal of this study is to propose a digital twin training and legal governance framework that can support lawful patrol autonomy, bounded arrest-support, and audit-ready validation for humanoid police robots in South Korea.

The central hypothesis is that a five-layer digital twin combining environment, robot, human-behavior, law-and-procedure, and evidence layers can make patrol autonomy more technically reproducible and legally auditable, while also demonstrating why arrest judgment, bodily restraint, and hazardous device activation must remain under explicit human police command. Because this is a design-science study rather than a completed field experiment, the hypothesis is treated as a design proposition to be operationalized through architecture, permission logic, and phase-gate benchmarks.

The paper addresses three research questions: RQ1, which patrol and arrest-support functions are suitable for different levels of autonomy; RQ2, how constitutional, criminal procedure, police equipment, privacy, AI governance, and compensation requirements can be translated into machine-readable controls; and RQ3, what benchmark metrics and phase-gate thresholds should govern movement from simulation to limited public pilots.

1.4. Contributions and Article Structure

This paper makes four contributions. First, it proposes a five-layer digital twin architecture composed of an environment twin, a robot twin, a human-behavior twin, a law-and-procedure twin, and an evidence twin. Second, it constructs an autonomy-permission framework that separates high-autonomy patrol from human-command arrest-support. Third, it defines a reproducible validation design built around comparator baselines, scenario families, metrics, and phase-gate thresholds rather than ad hoc pilot impressions. Fourth, it translates Korean constitutional, criminal procedure, police equipment, privacy, AI governance, and compensation law into implementable design controls. The remainder of the paper follows the requested structure: related work and legal background, materials and methods, findings, discussion and practical implications, conclusions, and limitations and future research.

2. RELATED WORK AND LEGAL BACKGROUND

2.1. Digital Twin and Humanoid Learning Research

Digital twins are not merely three-dimensional visual replicas. They are dynamic models linked to physical entities through data, simulation, prediction, and feedback loops [1]-[5]. In robotics, their value lies in enabling repeated training under variable weather, friction, lighting, crowd density, sensor noise, and communication latency without exposing humans to real danger. Large-scale parallel simulation further makes it possible to compress months of physical learning into hours of accelerated training [4]-[10].

The humanoid learning literature provides four especially important lessons for police applications. First, robust locomotion benefits from large-scale reinforcement learning in diversified simulation environments [6], [7]. Second, human skill transfer through teleoperation and imitation is crucial when behavior must reflect tacit field expertise, such as socially appropriate posture, safe interpersonal distance, or tactical positioning [8]. Third, sim-to-real transfer must be treated as a continuous calibration problem rather than a one-time export of a policy [9], [10]. Fourth, safe deployment in sensitive environments requires the twin to

represent not only physical states but also operational rules, intervention thresholds, and failure recovery logic. Table 1 synthesizes the main implications of the prior literature.

Table 1: Prior Research And Its Technical-Legal Implications

Domain	Representative Studies	Main Contribution	Implication For This Study
Digital twins	[1]-[5]	Define digital twins as dynamic, data-linked models rather than static visual scenes.	A policing twin must include interaction, procedures, and logs, not only geometry.
Humanoid locomotion	[6], [7]	Show that parallel simulation and robust reinforcement learning can produce stable whole-body control.	Patrol walking, balance, and obstacle avoidance should be learned in simulation first.
Teleoperation and imitation	[8]	Demonstrate that human motion and embodied skill can be transferred to humanoid policies.	Officer demonstrations can seed posture, spacing, escort, and equipment-handover behaviors.
Dynamic sim-to-real transfer	[9], [10]	Reduce the transfer gap through synchronization and continuous calibration.	Field logs must be fed back into the twin to maintain policy reliability.
Police robotics and law	[11]-[18]	Identify accountability, bias, force, surveillance, and legitimacy as central constraints.	Human command and auditability must be built into the learning architecture from the start.

2.2. Police Robotics and Public Safety Literature

International policing research has long considered robotics for surveillance, bomb disposal, search and rescue, patrol support, and crowd management [11]. More recent scholarship emphasizes that police robots differ from ordinary service robots because they mediate state power. Joh, Sankar, and Asaro argue that when robots become involved in search, seizure, or force, the analysis must center on constitutional rights, accountability, and the allocation of legal responsibility [12]-[14]. Hohensinn additionally shows that AI-supported police violence may reduce public blame directed at human officers, creating a moral distance that can normalize excessive intervention [15]. The Policing Project therefore recommends strong ex ante rules on force, transparency, and democratic accountability [16].

Korean studies reinforce the same lesson from a domestic perspective. Roh and Choi examine the technical feasibility and legal issues surrounding robot police and argue that deployment requires both capability certification and statutory interpretation [17]. Choi and Roh, in their study of real-time online patrol robots for cybercrime prevention, show that police robotics should be understood not only as a performance tool but also as a regulatory object subject to limits on monitoring scope, public notice, minimum collection, and legal authorization [18]. Although that study addresses cyberspace, its governance logic is highly relevant to physical police robots as well.

2.3. Applicable Legal Framework

The constitutional starting point is Article 12 of the Constitution of the Republic of Korea, which protects bodily liberty and due process [19]. The Criminal Procedure Act then specifies the legal structure of arrest, including arrest by warrant under Article 200-2 and warrantless arrest of a flagrant offender under Article 212 [20]. These provisions are written on the assumption that the actor who evaluates,

explains, records, and bears responsibility is a human officer or person. There is no explicit statutory basis for treating a humanoid robot as a subject of arrest authority.

The Act on the Performance of Duties by Police Officers is equally important. Article 10 permits the use of police equipment, requires safety education and inspection for hazardous police equipment, and prohibits arbitrary modification or abnormal use that harms life or body [21]. This suggests that a police robot is best understood, under current law, as a form of police equipment rather than an autonomous legal actor. At the same time, the robot's autonomy and contact capability create risks beyond those of conventional cameras, vehicles, or shields, which means that ordinary equipment rules are not sufficient on their own.

Privacy law adds a separate layer of constraints. The Personal Information Protection Act restricts the operation of mobile video processing devices in public places and, together with its Enforcement Decree, requires that data subjects can readily recognize that recording is taking place through lights, sounds, signs, broadcasts, or equivalent measures [22], [23]. The Personal Information Protection Commission has also issued specific guidance for mobile video processing devices, including robots and autonomous movers [24]. When training data are assembled, additional obligations arise under the State Compensation Act, the Product Liability Act, the Framework Act on Artificial Intelligence and its Enforcement Decree, the Intelligent Robots Development and Distribution Promotion Act, and the comparative benchmark established by the EU AI Act [25]-[30]. In the AI lifecycle context, the 2025 PIPC guide on generative AI further supports data minimization, lifecycle-based controls, and privacy-by-design governance [31].

3. MATERIALS AND METHODS

3.1. Research Design

The study adopts an interdisciplinary design-science method. Design-science research is appropriate because the object of analysis is not only an existing legal rule or an existing robot prototype, but a proposed socio-technical artifact: a deployable training and governance architecture for police humanoids. The method therefore combines technical synthesis, legal-doctrinal analysis, and architecture design.

The study does not claim to report field performance statistics. Instead, it develops artifacts that can be tested in later pilots: a layered digital twin, a permission matrix, a benchmark design, and a phase-gate operating procedure. This distinction is important because premature field claims would be misleading for a rights-sensitive police technology.

3.2. Materials

The materials consist of four groups. The first group is technical literature on digital twins, humanoid locomotion, reinforcement learning, imitation learning, teleoperation, and sim-to-real transfer [1]-[10]. The second group is police robotics and public-safety governance literature addressing search, seizure, force, accountability, bias, and legitimacy [11]-[18]. The third group is Korean legal and regulatory material, including constitutional, criminal procedure, police equipment, privacy, compensation, AI governance, and intelligent robot laws and guidance [19]-[29], [31]. The fourth group is comparative governance material, especially the EU AI Act as a benchmark for high-risk and prohibited law-enforcement AI design [30].

3.3. Analytical Procedure

The analysis proceeded in six steps. First, police-humanoid functions were decomposed into patrol, shared-autonomy patrol, arrest-support, and prohibited or highly restricted coercive functions. Second, the technical literature was mapped to the functional needs of locomotion, social navigation, perception, communication, sim-to-real transfer, and failure recovery. Third, the legal materials were analyzed doctrinally to extract authority limits, notice duties, data-processing constraints, equipment safety duties, liability pathways, and human-review requirements. Fourth, these requirements were translated into design controls such as reward constraints, approval states, geofencing, logging rules, emergency-stop requirements, and retention logic. Fifth, the resulting artifacts were organized into baseline comparators, scenario families, metrics, and phase-gate thresholds. Sixth, in-text citations and the reference list were reviewed for alignment with the revised argument.

3.4. Validity and Reproducibility Strategy

Because the paper is a conceptual design-and-governance study, validity is addressed through traceability rather than statistical inference. Each proposed control is

traceable to either a technical deployment risk or a legal-governance requirement. Reproducibility is addressed by specifying scenario families, comparator baselines, common log schemas, and go/no-go thresholds that later researchers can implement in simulation, hardware-in-the-loop testing, sandbox pilots, and limited public deployment.

Table 2: Methodological Components, Inputs, And Outputs

Method Component	Input Material	Analytical Task	Output
Technical synthesis	Digital twin, humanoid learning, teleoperation, and sim-to-real literature [1]-[10]	Identify trainable robot functions and failure modes	Robot, environment, human-behavior, and validation requirements
Legal-doctrinal analysis	Korean constitutional, criminal procedure, police, privacy, AI, compensation, and robot laws [19]-[29], [31]	Extract authority, privacy, safety, liability, and review constraints	Law-and-procedure controls and prohibited-function boundaries
Comparative governance mapping	Police robotics governance and EU AI Act references [11]-[18], [30]	Translate high-risk AI and democratic accountability principles	Human-in-command, audit, and bias-review requirements
Design synthesis	Function decomposition and operational assumptions	Integrate technical and legal constraints into one architecture	Five-layer digital twin and autonomy-permission matrix
Validation design	Scenario families, baselines, and phase-gate logic	Define reproducible evaluation criteria	Benchmark matrix and deployment SOP

4. FINDINGS: DIGITAL TWIN TRAINING AND GOVERNANCE FRAMEWORK

The findings are presented as design-science outputs rather than empirical trial results. They answer the research questions by specifying what the architecture should contain, how autonomy should be bounded, and how later pilots should be evaluated.

4.1. Finding 1: Five-Layer Digital Twin Architecture

The first finding is that a policing digital twin must include five linked layers. The environment twin models roads, stations, airports, corridors, stairs, doors, lighting, weather, floor friction, crowd density, and zones of access. The robot twin models degrees of freedom, torque and speed limits, sensors, battery, network delay, fault states, and emergency-stop logic. The human-behavior twin models pedestrians, vulnerable persons, suspects, bystanders, and human officers. The law-and-procedure twin encodes warnings, approval states, permissible contact, privacy notices, retention rules, and escalation thresholds. The evidence twin stores synchronized logs, hashes, timestamps, operator interventions, model versions, and replay packages.

The principal value of this structure is that technical performance and legal validity can be trained together. A policy that completes a patrol route but violates notice rules, continues to track a person without authorization, or initiates bodily contact without explicit approval should not receive a high reward. Conversely, a policy that sacrifices speed to maintain safe distance, report uncertainty, and escalate to a human officer may be operationally superior in public safety. In this sense, the digital twin becomes not only a simulator but also a compliance and evidence architecture.

Table 3: Five-Layer Digital Twin Structure For Police Humanoid Training

Twin Layer	Main Elements	Training Objective	Key Legal / Procedural Design Point
Environment twin	GIS / BIM maps, lighting, weather, stairs, doors, crowd flow, public-semi-public-private zones	Route following, obstacle avoidance, spatial awareness	Zone classification and geofencing must reflect lawful patrol boundaries.
Robot twin	Kinematics, torque and speed limits, battery, sensors, communication delay, emergency stop	Stable locomotion, fault recovery, safe control	Hard motion limits and dual-stop logic are mandatory.
Human-behavior twin	Pedestrians, officers, suspects, vulnerable persons, bystanders	Social navigation, de-escalation, separation, cooperation	Policies must protect vulnerable persons and avoid discriminatory targeting.
Law-and-procedure twin	Warnings, approval states, contact permissions, retention rules, notice signals	Only lawful action sequences receive reward	No bodily restraint is permissible before explicit human approval.
Evidence twin	Synchronized logs, hashes, timestamps, model versions, operator interventions	Replay, accountability, post-incident audit	Evidentiary integrity and chain-of-custody logic must be preserved.

4.2. Finding 2: Patrol Learning Can Support Bounded High Autonomy

Patrol functions should be decomposed into legally lower-risk tasks that can tolerate a higher level of autonomy. These include route following, stair climbing, obstacle avoidance, facility inspection, anomaly detection, warning broadcast, remote presence, scene preservation, and autonomous return to a charging or docking point. The learning target is not simply to maximize path completion. It is to maximize lawful, non-threatening, and explainable patrol behavior under variable environmental conditions.

Reward design for patrol therefore includes positive terms for route completion, no-fall operation, timely reporting, and safe interaction, and negative terms for collisions, unnecessary pursuit, excessive proximity to crowds, notice failure, recording in prohibited zones, overconfident anomaly classification, and risky continuation under sensor or network uncertainty. In dense public environments, non-threatening presence must be treated as a technical objective. The robot should not startle crowds, block natural human flow, or behave like an opaque enforcement machine.

4.3. Finding 3: Arrest-Support Requires Human-in-Command Permission Logic

Arrest-related learning must be handled under a much stricter model than patrol. This paper divides it into six stages: approach, containment, approval wait, support intervention, scene preservation, and transfer assistance. The key point is that the robot is trained as an arrest-support platform rather than an independent arrest actor. Before human authorization, the robot may position itself to open an approach corridor for officers, illuminate an area, mark exits, create a buffer from bystanders, and send synchronized video. After explicit approval, it may assist with equipment handover, doorway control, fall prevention, bystander separation, and evidentiary recording.

By contrast, direct bodily restraint, limb control, cuffing, neck or thoracic pressure, weapon activation, and other contact-rich coercive acts should not be part of ordinary autonomous deployment. Even in research settings, these behaviors require hard constraints

on torque, speed, contact pressure, contact location, and duration. When uncertainty exceeds a threshold, the system should automatically retreat or stop rather than improvise. In other words, arrest-support learning is not about maximizing capture success at any cost. It is about ensuring that legally and physically forbidden boundaries are not crossed.

Table 4: Autonomy Levels, Permissible Functions, And Legal Assessment

Autonomy Level	Representative Functions	Expected Autonomy	Legal Assessment
Level A: Autonomous patrol	Route patrol, obstacle avoidance, facility inspection, SOS relay, auto return	High	Generally permissible when notice, safety certification, and logging are ensured.
Level B: Shared-autonomy patrol	Anomaly detection, warning broadcast, scene preservation, remote presence	Medium to high under supervision	Permissible when uncertainty is visible and human review is available.
Level C: Arrest-support	Corridor securing, illumination, equipment delivery, bystander separation, door control, video record	Human-in-command required	Permissible only after explicit human approval and within a narrow support role.
Level D: Bodily restraint or hazardous device use	Limb control, cuffing, direct physical restraint, weaponized actuation	Not suitable for current autonomous deployment	Requires explicit legislation and stricter certification; should remain prohibited in current practice.

4.4. Finding 4: Sim-to-Real Transfer Must Include Evidence Preservation

A simulated policy should never be transferred to the field in a single step. The validation path should include adversarial software testing, hardware-in-the-loop verification, indoor controlled trials, a restricted sandbox pilot, and only then limited public deployment. Stress scenarios must include sudden crowd inflow, low-light conditions, wet surfaces, communication delay, sensor dropout, partial actuator failure, confusing human gestures, and conflicting operator commands. The goal is not only to improve nominal performance but also to verify graceful degradation and safe failure.

Evaluation metrics must combine technical and procedural indicators. Patrol evaluation should include path-completion rate, mean time without fall, anomaly false-positive rate, response latency, and complaint rate. Arrest-support evaluation should include zero pre-approval contact events, zero pressure-threshold exceedance events, 100 percent emergency-stop success, 100 percent log integrity, and zero missing original records. For policing, the evidence package is part of the system itself: sensor inputs, model version, operator interventions, approval signals, warnings, and stop events must be bundled into a replayable event file that supports audit and judicial review.

4.5. Finding 5: Benchmark Design Requires Baselines And Tail-Risk Metrics

A validation framework is meaningful only if it can distinguish the value of the proposed architecture from simpler alternatives. The paper therefore defines four comparator baselines for later experiments: B0, teleoperation-only operation without learned autonomy; B1, a patrol robot trained for navigation and anomaly detection but without humanoid whole-body capability; B2, a humanoid twin trained for technical mission performance without law-and-procedure or evidence layers; and B3, the proposed integrated five-layer architecture. Comparison across these baselines isolates the added value of humanoid embodiment, law-in-the-loop constraints, and audit-oriented logging.

Scenario families should include open public-facility patrol, dense transit-hub patrol, low-light and wet-surface disturbance, arrest-support corridor control, scene preservation, and communication-loss recovery. Each scenario should be executed under randomized crowd, weather, lighting, sensor-noise, and network-delay conditions using pre-registered seeds and common log schemas. Reported outcomes should include both mean performance and worst-case tail events, because policing risk is dominated by low-frequency high-impact failures rather than average performance alone.

5. DISCUSSION

5.1. Interpretation And Justification Of The Findings

The findings support the central design proposition in two ways. Technically, a digital twin can expose a humanoid policy to rare but important conditions that would be unsafe or impractical to stage repeatedly in public. Legally, the same twin can make authority limits visible by encoding approval states, recording boundaries, prohibited contact, retention rules, and evidence logging into the training objective. The framework therefore reduces the risk that legality will be treated as a checklist after a technically optimized robot has already been built.

The distinction between patrol and arrest-support is also justified. Patrol tasks can often be evaluated through route completion, safety margins, notice compliance, anomaly reporting, and operator override. Arrest-related tasks are different because they can directly affect bodily liberty and because a robot cannot bear the duties, explanations, judgment, or responsibility that current criminal procedure assumes of a human actor. The appropriate design conclusion is therefore asymmetric: patrol can move toward bounded high autonomy, while arrest judgment and coercive contact should remain under human police command.

5.2. Legal And Governance Implications

The most fundamental legal question is who may decide and initiate arrest. Article 12 of the Constitution requires due process for any restriction of bodily liberty [19]. The Criminal Procedure Act structures arrest around human actors who can interpret facts, give reasons, hear objections, prepare records, and bear legal responsibility [20]. Although Article 212 states that a flagrant offender may be arrested without a warrant by any person, extending that phrase to an AI robot is legally fragile. A robot has no legal personhood, cannot meaningfully explain reasons for arrest, cannot independently bear procedural duties after arrest, and cannot itself be sanctioned for unlawful restraint. The safer legal interpretation is therefore to treat the robot as arrest-support equipment, not as the subject of arrest authority.

Article 10 of the Act on the Performance of Duties by Police Officers permits the use of police equipment, requires safety education and inspection for hazardous equipment, and prohibits arbitrary modification or abnormal use that endangers life or body [21]. A humanoid robot equipped with actuators, mobility, contact surfaces, and communication links can therefore be interpreted as police equipment, but it is not equivalent to ordinary cameras or vehicles. Once the robot can block paths, manipulate objects, or physically interact with people, the risk profile changes substantially. A separate category for autonomous or semi-autonomous police robotic equipment should therefore include torque ceilings, speed limits, geofencing, approved contact areas, default deactivation of hazardous functions, dual emergency-stop devices, model-version registration, and periodic recertification.

A patrol humanoid that records public space is functionally a mobile video processing device. Under the Personal Information Protection Act and its Enforcement Decree, recording in public places is restricted unless there is an applicable legal basis or clearly recognizable notice, and the recording fact must be made easy for data subjects to perceive through lights, sounds, signs, broadcasts, or equivalent means [22], [23]. Operationally, patrol zones should be divided into public, semi-public, and non-public spaces. Open roads, station halls, and airport public areas can be patrolled with explicit notice and a documented legal basis. Semi-public spaces, such as apartment common areas or shopping malls, require operator consent and supplementary notice. Private offices, residential interiors, hotel rooms, or back-of-store spaces should be excluded absent a separate legal basis or warrant.

Liability and redress also require design attention. If a police robot malfunctions or misclassifies a person and causes harm, the victim should not carry the burden of untangling a complex technical chain before receiving relief.

Because the robot is deployed in official policing, primary public liability should first be considered under the State Compensation Act [25]. Where harm is attributable to design defects, software defects, defective updates, or maintenance failure, product liability and contract-based allocation remain relevant [26]. A realistic framework distributes liability across the operating agency, the responsible officer, the manufacturer, and the maintenance contractor, but it must preserve original video, derived analytics, model versions, operator interventions, approval states, and emergency-stop history to make that distribution possible.

Finally, police patrol and arrest-support robots are likely to be treated as high-impact AI because they can affect bodily safety, freedom, and other basic rights [27], [28]. The Intelligent Robots Development and Distribution Promotion Act supplies a domestic policy backdrop for responsible robot development and distribution [29]. Comparative law is also instructive: the EU AI Act treats several law-enforcement AI uses, especially biometric and other fundamental-rights-sensitive uses, as prohibited or high-risk [30]. A deployable police humanoid should therefore undergo demographic stress testing, threshold calibration, false-positive analysis, model-card documentation, and third-party review. Citizens and officers alike must understand that risk scores and annotations are analytical aids, not substitutes for legal judgment.

Table 5: Legal Issues, Risks, And Procedural Controls

Issue	Main Risk	Procedural Solution	Control Point
Arrest authority and due process	Unlawful bodily restraint by a non-human decision-maker	Keep arrest decision and bodily restraint under explicit human command	No contact before approval; automatic retreat on uncertainty
Police equipment and use of force	Excessive force, unsafe modification, uncontrolled hazardous capability	Create a separate certification regime for police robots and disable hazardous functions by default	Torque ceilings, geofencing, dual-stop devices, recurrent inspection
Privacy and recording	Invisible recording, over-collection, illegal capture in restricted zones	Classify zones, provide notice, minimize retention, prioritize synthetic data	Lights, signs, broadcasts, access control, deletion rules
Liability and redress	Victim faces delayed or fragmented compensation	State-first compensation with downstream recourse to manufacturer or maintainer	Immutable logs, insurance or dedicated relief fund
High-impact AI and bias	Opaque risk scoring and discriminatory targeting	External audit, model documentation, demographic stress tests, human review	Threshold review, replay, explanation, recertification

5.3. Practical Implications

The practical implications are concrete. Police agencies should begin with low-risk patrol functions, visible notice, remote supervision, and explicit prohibited-function lists rather than attempting broad autonomous enforcement. Vendors should design the robot as an auditable public-safety system, not merely as a locomotion platform with cameras. Regulators should require separate certification for police robotic equipment and should treat law-in-the-loop and audit-by-design controls as minimum deployment conditions. Researchers should report baselines, scenario seeds, log schemas, and worst-case failures so that pilots can be compared across institutions.

For police managers, the framework also clarifies procurement and training decisions. A public agency should not simply ask whether a humanoid robot can patrol a hallway or climb stairs. It should ask whether the system can prove

notice compliance, prevent unauthorized contact, preserve chain-of-custody evidence, recover safely from network failure, and show a complete record of human approvals and overrides. These requirements can be placed in procurement specifications, sandbox protocols, officer training manuals, and post-incident review procedures.

Table 6: Practical Implications And Implementation Responsibilities

Stakeholder	Practical Action	Reason	Expected Output
Police agency	Define permitted and prohibited functions before procurement	Prevents technology-led expansion of police authority	Function map and approval-state policy
Robot vendor	Build law-and-procedure and evidence layers into the product architecture	Makes legal compliance technically enforceable	Audit-ready software and log package
Regulator / auditor	Require phase-gate testing, bias review, and recertification	Limits unsafe scale-up after narrow prototype success	Certification checklist and external audit report
Human operator	Use a live interface showing uncertainty, recording status, approval status, and emergency-stop readiness	Turns oversight from nominal monitoring into effective control	Human-in-command operating record
Research community	Publish baselines, scenario seeds, and failure cases	Improves reproducibility and avoids impressionistic pilots	Comparable benchmark results

5.4. Phased Deployment And Standard Operating Procedure

A realistic implementation path should proceed through seven phases rather than direct field deployment. Phase 1 is functional definition and legal mapping. The agency must define which functions are included, which are prohibited, and what legal basis applies to each. Phase 2 is data governance and twin construction, including synthetic-data prioritization, pseudonymization, zone mapping, and event-log design. Phase 3 is simulation training, where reinforcement learning, imitation learning, and scenario stress testing are conducted in the multi-layer twin.

Phase 4 is hardware-in-the-loop and indoor verification. Here, real actuators, sensors, and emergency-stop devices are tested against edge cases and fail-safe conditions. Phase 5 is a regulatory sandbox or restricted pilot under remote human supervision in a tightly bounded area. Phase 6 is limited public deployment in selected open spaces with visible notices and live monitoring. Phase 7 is routine deployment with periodic recertification, incident review, and external audit. At every transition, technical performance must be judged together with legal and procedural compliance.

The sandbox stage is especially important because it converts an engineering prototype into a public-administration instrument. During this stage, the supervisory interface should display the robot's current state, uncertainty score, recording status, approval status, network health, and emergency-stop readiness in real time. A multi-agency governance structure involving the police, privacy regulator, AI regulator, and legal experts is desirable, and an independent audit board should review safety incidents, bias indicators, complaint trends, and model updates before any scale-up decision is made.

Table 7: Benchmark Matrix And Phase-Gate Validation Thresholds

Phase	Validation Method	Key Metrics	Go / No-Go Threshold
1. Legal mapping and prohibited-function definition	Statute mapping, role allocation review, prohibited-act matrix	Function-to-legal-basis coverage; prohibited-function completeness	100% of deployed functions mapped; all prohibited acts encoded as hard constraints
2. Twin construction and data governance	Zone classification, synthetic-data prioritization, log-schema design, privacy impact review	Notice-rule coverage; synthetic-data ratio; log-field completeness	Privacy impact assessment approved; notice zones defined; complete log schema
3. Simulation training	RL / imitation learning under randomized crowd, lighting, weather, and noise	Path completion; fall rate; anomaly precision/recall; unauthorized contact events	Mission metrics meet target; unauthorized contact = 0
4. HIL and indoor verification	Real actuators, sensors, pressure limits, emergency stop, network-loss tests	Emergency-stop success; safe-retreat success; torque/pressure exceedance count	Emergency stop = 100%; safe retreat = 100%; exceedances = 0
5. Sandbox pilot	Restricted-area supervised pilot with live operator interface	Complaint count; override latency; false-alarm rate; evidence completeness	No severe safety event; approved latency met; evidence completeness = 100%
6. Limited public deployment	Selected open-space patrol with visible notice and continuous monitoring	Notice compliance; false-positive rate; retention/deletion compliance; review rate	Notice compliance = 100%; retention/deletion compliance = 100%; false positives within approved threshold
7. Routine deployment and recertification	Periodic audit, bias review, version control, incident feedback	Audit pass rate; bias drift; model traceability; patch compliance	Independent audit passed; no unexplained bias drift; full traceability

5.5. Difference From Prior Work And Overall Impact

The present study differs from prior work in its integration logic. Existing technical studies generally improve locomotion, teleoperation, imitation learning, or sim-to-real transfer [6]-[10]. Existing legal scholarship tends to analyze police surveillance, force, constitutional rights, or accountability [12]-[18]. Pilot-oriented public-safety reports identify promising use cases but rarely connect them to benchmark design and certification thresholds [11], [16]-[18]. This paper treats deployment as a single socio-technical problem in which training architecture, legal constraints, evidence preservation, and certification criteria are co-designed.

The overall impact of the framework is that it can help future pilots move beyond proof-of-concept demonstrations. A pilot that merely shows that a humanoid can walk a route is not sufficient for policing. A meaningful pilot must also show that the robot respects permission boundaries, records lawfully, escalates uncertainty to humans, preserves evidence integrity, and fails safely. The proposed law-and-procedure twin, evidence twin, and phase-gate matrix provide a path toward that more rigorous standard.

6. CONCLUSIONS

This paper proposed a digital twin-based training and legal governance framework for humanoid police robots in patrol and arrest-

support operations. The main conclusion is that digital twins offer the most realistic route for preparing police humanoids because they allow locomotion, perception, interaction, and operational decision sequences to be tested without exposing citizens to uncontrolled trial and error. Yet for public safety, a digital twin must do more than reproduce physical scenes. It must also reproduce authority limits, approval pathways, privacy requirements, and evidence-preservation logic.

Under the current Korean legal framework, a humanoid robot can plausibly evolve into an advanced police equipment platform for patrol, observation, warning, remote presence, scene preservation, and narrowly defined arrest-support functions. It should not, however, be treated as an independent subject of arrest authority or autonomous coercive force. The most defensible near-term model is therefore a digitally trained police-support humanoid operating under explicit human command, with law-in-the-loop constraints, privacy-by-design controls, audit-ready logs, and phase-gate validation thresholds.

7. LIMITATIONS AND FUTURE WORK

This paper is a conceptual design-and-governance study and does not report a field-deployed humanoid prototype or completed comparative benchmark results. The proposed validation metrics, baselines, and operating procedure therefore remain to be tested through controlled pilots. The legal analysis is also centered on South Korea, although comparative references are used to strengthen the governance design.

Future research should build a Korean patrol-environment twin, implement the proposed scenario families, and compare B0 to B3 baselines under repeated randomized conditions. Empirical studies should measure complaint rates, false positives, response latency, operator workload, emergency-stop reliability, safe-retreat performance, and public acceptance. Further work should also examine demographic bias, privacy notices in crowded settings, evidentiary admissibility of robot logs, and comparative constitutional and administrative-law treatment of police robots across jurisdictions. The validation matrix introduced here should therefore be read as a pre-registered evaluation design rather than finished performance evidence.

REFERENCES

[1] M. Grieves and J. Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," in *Transdisciplinary Perspectives on Complex Systems*, F.-J. Kahlen, S. Flumerfelt and A. Alves, Eds., Springer, 2017, pp. 85-113.

[2] A. Rasheed, O. San and T. Kvamsdal, "Digital Twins: State of the Art Theory and Practice, Challenges, and Open Research Questions," *Journal of Industrial Information Integration*, Vol. 30, 2022, Article 100383.

[3] A. Thelen, X. Zhang, O. Fink, Y. Lu, S. Ghosh, B. D. Youn, M. D. Todd, S. Mahadevan, C. Hu and Z. Hu, "A Comprehensive Review of Digital Twin-Part 1: Modeling and Twinning Enabling Technologies," arXiv:2208.14197, 2022.

[4] J. Li and S. X. Yang, "Digital Twins to Embodied Artificial Intelligence: Review and Perspective," *Intelligence & Robotics*, Vol. 5, No. 1, 2025, pp. 202-227.

[5] J. F. Yao et al., "Systematic Review of Digital Twin Technology and Applications," Visual Computing for Industry, *Biomedicine*, and Art, Vol. 6, 2023, Article 10.

[6] X. Gu, Y.-J. Wang and J. Chen, "Humanoid-Gym: Reinforcement Learning for Humanoid Robot with Zero-Shot Sim2Real Transfer," arXiv:2404.05695, 2024.

[7] D. Ferigo, G. Camoriano, S. Traversaro and F. Nori, "On the Emergence of Whole-Body Strategies from Humanoid Robot Push-Recovery Learning," *IEEE Robotics and Automation Letters*, Vol. 6, No. 4, 2021, pp. 8561-8568.

[8] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu and G. Shi, "Learning Human-to-Humanoid Real-Time Whole-Body Teleoperation," arXiv:2403.04436, 2024.

[9] J. Abou-Chakra, L. Sun, K. Rana, B. May, K. Schmeckpeper, M. V. Minniti and L. Herlant, "Real-is-Sim: Bridging the Sim-to-Real Gap with a Dynamic Digital Twin for Real-World Robot Policy Evaluation," arXiv:2504.03597, 2025.

[10] Y. Liu, H. Xu, D. Liu and L. Wang, "A Digital Twin-Based Sim-to-Real Transfer for Deep Reinforcement Learning-Enabled Industrial Robot Grasping," *Robotics and Computer-Integrated Manufacturing*, Vol. 75, 2022, Article 102287.

[11] United Nations Interregional Crime and Justice Research Institute and INTERPOL, *Artificial Intelligence and Robotics for Law Enforcement*, UNICRI, 2019.

[12] E. E. Joh, "Policing Police Robots," *UCLA Law Review Discourse*, Vol. 64, 2016, pp. 516-543.

[13] V. Sankar, "What Happens When Police Robots Violate the Constitution?," Vanderbilt *Journal of Entertainment & Technology Law*, Vol. 20, No. 3, 2018, pp. 947-986.

[14] P. M. Asaro, "Automating Police Use of Force," Proceedings of the AAAI Spring Symposium on *Ethically Bounded Artificial Intelligence*, 2015.

[15] L. Hohensinn, "Who Guards the Guards with AI-Driven Robots? Perceived Ethicalness of Police Violence in the Age of AI," *Public Management Review,* 2024.

[16] The Policing Project, Police Robots: A Policy Framework, *New York University School of Law*, 2023.

[17] S. Roh and W. Choi, "A Study on the Technical Feasibility and Legal Issues for the Introduction of Robot Police," *Asia-Pacific Journal of Convergent Research Interchange*, Vol. 10, No. 3, 2024, pp. 411-426.

[18] W. Choi and S. Roh, "Development and Legal Review of Real-Time Online Patrol Robots for Cybercrime Prevention and Suppression," *Hannam Journal of Law & Technology*, Vol. 31, No. 1, 2025, pp. 53-83. DOI: 10.32430/ilst.2025.31.1.53.

[19] Constitution of the Republic of Korea, *Constitutional Law* No. 10, 1987, National Law Information Center.

[20] Criminal Procedure Act, Act No. 20796, 2025, National Law Information Center.

[21] Act on the Performance of Duties by Police Officers, Act No. 21014, 2025, National Law Information Center.

[22] Personal Information Protection Act, Act No. 20897, 2025, National Law Information Center.

[23] Enforcement Decree of the Personal Information Protection Act, Presidential Decree No. 35780, 2025, National Law Information Center.

[24] Personal Information Protection Commission, Guide on Personal Image Information Protection and Use for Mobile Video Information Processing Devices, 2024.

[25] State Compensation Act, Act No. 20635, 2025, National Law Information Center.

[26] Product Liability Act, Act No. 14764, 2018, National Law Information Center.

[27] Framework Act on the Development of Artificial Intelligence and the Creation of a Foundation of Trust, Act No. 20676, 2025 (effective 2026), National Law Information Center.

[28] Enforcement Decree of the Framework Act on the Development of Artificial Intelligence and the Creation of a Foundation of Trust, Presidential Decree No. 36053, 2026, National Law Information Center.

[29] Intelligent Robots Development and Distribution Promotion Act, Act No. 19412, 2023, National Law Information Center.

[30] 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 (Artificial Intelligence Act), Official Journal of the European Union, 2024.

[31] Personal Information Protection Commission, Guide to Personal Information Processing for the Development and Use of Generative AI, 2025.