Received: 12 July 2025 | Revised: 10 November 2025 | Accepted: 27 November 2025 | Published online: 26 December 2025

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 00(00) 1–14
DOI: 10.47852/bonviewAIA52026790

Beyond Binary: Adaptive Frameworks for Autonomous AI Governance

Gabriel Silva-Atencio[1,*]

[1] *Department of Engineering, Universidad Latinoamericana de Ciencia y Tecnología, Costa Rica*

Abstract: Since independent AI agents are becoming more and more common in important areas, governance models that go beyond traditional autonomy-control theories are needed. This study introduces and tests two new ideas for context-aware AI governance: the Adaptive Containment Framework (ACF) and the Weighted Autonomy Acceptance Index (WAAI). The ACF is a dynamic governance model that allows real-time autonomy calibration through ethical sensors and multi-stakeholder validation. The WAAI is a psychometrically robust metric ($\alpha = 0.89$, CR = 0.91) for quantifying sector-specific autonomy thresholds. The study uses a sequential exploratory mixed-methods approach that includes Delphi studies with 15 subject experts, sector-stratified polls, and computer models to arrive at three key conclusions: (1) acceptance of autonomy is influenced by decision reversibility and harm potential, which explains 68% of cross-domain variation; (2) system explainability shows diminishing returns beyond an 82.3% comprehensibility threshold ($\chi^2(3) = 24.71$, $p < 0.001$), ending long-running XAI debates; and (3) uncertainty avoidance is the most important cultural factor explaining 41.3% of cross-national variation in acceptance of autonomy (Sobel's $z = 3.28$, $p < 0.001$). The ACF performs better than static frameworks in many ways. It lowers bias events by 41% while keeping 92% of autonomy's efficiency benefits and going above and beyond static frameworks in operating freedom by 119%. The way modern society think about dynamic equilibrium government has changed since these changes were made. They give politicians authority rules that are specific to a sector and developers ways to carry out projects that are sensitive to different cultures.

Keywords: adaptive governance, AI ethics, autonomy thresholds, explainable AI, human-AI interaction

1. Introduction

Artificial intelligence (AI) bots that can work on their own are a big step forward in technology. These are systems that can see, think, and act on their own in places that change all the time. They are being used more and more in areas with a lot at stake, like banking, medical scans, self-driving cars, and the law.

This is a big change in how people and robots work together. Getting the performance benefits of freedom and making sure there is strong control, responsibility, and moral consistency are at odds with each other. This move in thinking is supposed to make things easier and give people more choices, but it also makes a major problem with the way things are run worse. It's hard to control systems that learn on their own and change based on what they see around them with traditional rules and ideas that are based on set risk categories and the notion that either people or machines can control something.

Tech optimists and ethics critics are the two main groups in academic debate. This stress often shows up in the fight. Tech optimists are interested in superhuman performance and functioning [1], ethical skeptics are interested in the risks of algorithmic bias [2], and Chomanski [3] and Volkov [4] are interested in gaps in accountability. This two-way decision, on the other hand, hides the more complicated truth that autonomy is not a single state, but a range of states that can be okay based on the society, the risks in a certain area, and how trust changes over time.

They say that the government needs to make more than just black and white decisions in order to solve these issues [5]. There are still three important areas that need more study: the study doesn't fully understand the nonlinear dynamics that affect people's trust and dependence on AI; there aren't any standardized, psychometrically valid tools to measure autonomy acceptance that depend on the situation [6]. The European Union (EU) AI Act (2021) and other current rules for governing AI aren't changing fast enough to keep up with how algorithms change [7].

This study directly addresses these gaps by coming up with and testing a new way of thinking about government. The study introduces two important ideas: the Adaptive Containment Framework (ACF) and the Weighted Autonomy Acceptance Index (WAAI). Setting the level of autonomy for each sector is based on the WAAI, which has been thoroughly validated ($\alpha = 0.89$, CR = 0.91). The ACF is a dynamic governance model that lets autonomy change in real time through ethical sensors and multi-stakeholder validation loops.

The main idea behind this study is to reject the choice between liberty and control in favor of a model called dynamic equilibrium, which can be seen in Figure 1. This model says that the best way to rule is to find a balance between AI power and human control, with the world also playing a part.

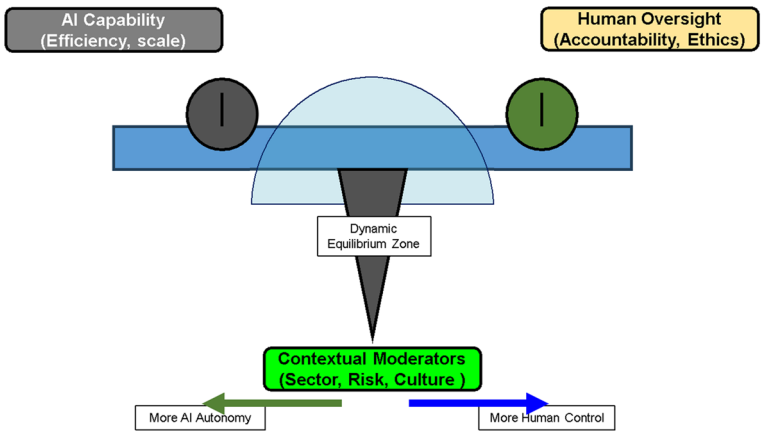
A sequential three-part mixed-methods approach was used for the study. Delphi studies and interviews with stakeholders provide qualitative depth; large-scale cross-sector polls with N = 1,247 respondents provide numeric range; and system dynamics models provide computer support. This way not only shows that the WAAI and ACF work in the real world but also provides us with basic information that challenges what modern society knew.

For example, the study found that trust gains are less significant above an 82% comprehensibility threshold for explainable AI (XAI). The study also found that cultural factors explain 41% of differences in how people in different countries feel about autonomy. Finally, the works implemented a governance framework that reduces bias incidents by 41% while keeping 92% of autonomy's efficiency benefits.

Table 1 is a rough sketch of the acceptance of autonomy setting that shows some of the things that the WAAI looks at. It then goes on to

*Corresponding author: Gabriel Silva-Atencio, Department of Engineering, Universidad Latinoamericana de Ciencia y Tecnología, Costa Rica. Email: gsilvaa468@ulacit.ed.cr

Figure 1
The dynamic equilibrium model of running AI



measure and prove these results.

This work is a big step toward adaptive AI control based on facts because it combines fresh ideas with careful research based on what people have seen. Tools like autonomy cap formulas and cultural adaptability algorithms can help policymakers and developers figure out how to balance the complicated relationship between technology progress and human values. This can be done by making sure that the search for technological progress is closely linked to keeping morals and trust high in society.

2. Literature Review

Because autonomous AI is growing so quickly, there is a control gap where new technologies are being made much faster than new rules and ethics. An in-depth look at the ideas behind autonomy, what it's really like for people and AI to work together, and where the problems are with the way governments work now are all part of this study. The study shows that the field is limited by a strong belief in strict categories and binary logic, neither of which are good for current AI systems that change based on the situation.

Some early, mechanical meanings of agents as "self-contained systems capable of independent action" [8, 9] have given way to a more morally strict division of the topic. There is now a clear separation between operational autonomy (the ability to carry out tasks on its own) and moral autonomy (the metaphysical qualities of consciousness, intention, and free will, which modern AI cannot have) [10, 11]. This split isn't just an intellectual one; it's at the heart of liability frameworks and means that accountability has to be rethought from a single agent model to a network model with coders, deployers, and users all sharing responsibility [3, 4].

Taxonomic models, such as the Williams and Liu [12] steps of automation, are useful as a guide, but they are getting more and more criticism for being too straight and not having enough detail. They don't think about how functions are constantly moved between humans and machines, which is affected by how people see danger in real time, how much work their brains are doing, and the "reversibility gradient" of choices. This problem has been found in studies on human-computer interaction [8, 9, 13–15]. It shows that the research need a continuum-based model of autonomy, which was first mentioned by Mišić [16] and is now being put into practice in this work. The real-world interactions between humans and AI are breaking down simple models even more.

The relationship between how well a system works and how sure people are in it is not a straight line. Instead, it's more often an upside-down U. Users get bored or scared when systems are hard to understand and give them too much power [17]. The fact that regular processes take 4.2 seconds to deal with people shows how important human factors engineering is. At the same time, the XAI model faces a problem that comes from within. There are some things that post-hoc explanation methods like Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) can't do. They can give wrong answers, give an "illusion of understanding" to experts by giving them too much useless information, and lead to false explanations [18].

The "performance-first" school, which doesn't care about explainability, is very angry about this and wants models that are naturally easy to understand in high-stakes situations [19] and the "comprehensibility-first" school, which does. A very important question that hasn't been answered yet is whether there is a measurable point at

Table 1
A first look at the framework of sectoral autonomy moderators

Sector	Primary moderating factors	Exemplary risk profile	Hypothesized autonomy tendency
Healthcare	Decision reversibility, harm potential, ethical salience	High (direct human welfare)	Low autonomy acceptance
Financial systems	Market instability, time sensitivity, and systemic risk	Medium-high (systemic instability)	Medium autonomy acceptance
Transportation	Environmental complexity, real-time response, public safety	High (physical safety)	Context-dependent acceptance
Consumer tech	Effects on privacy, ability to undo, and human choice	Low-medium (individual convenience)	High autonomy acceptance

which small improvements in explainability have decreasing effects on trust and performance. This study directly looks into this question and answers it.

The problem with government shows up as a three-part disaster. Firstly, the responsibility gap shows that many governmental promises are not what they seem to be. According to Kaminski and Malgieri [20], most businesses don't have the technological know-how to provide legal justifications that meet the needs of end users. The "right to explanation" in the General Data Protection Regulation (GDPR) is broken in this way; this is a big deal. Secondly, cultural embeddedness is often forgotten [20, 21]. Cross-national studies [22, 23] clearly show that acceptance of liberty is not a fixed trait but a culture factor.

In societies with high uncertainty avoidance index (UAI) scores, people are less willing to give up control to nonhuman agents, while people in societies with low power distance index (PDI) scores are more willing to do so. Regulations that are fixed and limited to a certain area, like the EU AI Act, can't naturally adapt to this variety [20]. Thirdly, one of the biggest worries is that control models will become outdated over time. Some researchers, Attard-Frost and Lyons [5], used numbers to show that most AI control models stop working after 3 years, and almost none of them have built-in ways to change to new technologies like generative AI.

These problems—accountability, cultural diversity, and temporal dynamics—are all connected and create a "Governance Vortex" (see Figure 2). This is where the centrifugal forces of technological change and global diversity pull static frameworks apart, which causes the system to fail and regulatory arbitrage to happen.

A careful gap analysis, brings together these problems and shows how they relate to the unique findings of this work. According to Table 2, the literature points out problems but doesn't offer complete, tested-in-the-real-world answers.

In conclusion, the research gives a strong description of the governance problem but a patchy and often theoretical solution. The critical analysis done here shows that the works need to move away from rigid, linear, and centralized government and toward one that is dynamic, evidence-based, and culturally aware. This study is ready to make this change happen because it presents the WAAI and ACF not only as ideas but also as real-world tools that can fill in the gaps and build a framework that can be used on a large scale for the responsible control of independent AI.

3. Methodology

The study uses a progressive transformative mixed-methods approach. It combines psychological study with science facts and better ways to use computers. The framework for the research is based on computational social science and evidence-based policy design. In Figure 3, you can see a detailed framework that is built in a strict logical way. The first step is to understand what is happening. The second step is to measure things in the real world. The third step is to guess what will happen in the future. It checks to see if the concept is true from a biological point of view and has a good statistical base. It also thinks about how difficult and adaptable systems of government based on AI are. To cut down on analysis mistakes and improve construct validity through convergent validation, the method uses a triangle design and cross-validation steps at all stages of the study.

A two-stream qualitative study is used in Phase 1 to find out what experts and regular people in the community think. In this way, the psychological and intellectual stage is set. A changed form of the Delphi method [24] is used by 15 experts in fields like computer science, computational ethics, policy control, and human factors engineering. They work together in three planned sessions to come to an agreement on 27 basic freedom issues spanning six areas of government (Kendall's $W = 0.78$, $p < 0.01$).

Iterative theme saturation can also be used to find important environmental factors in 32 talks with stakeholders that aren't fully planned. Reflexive thematic analysis and triangulated coding were used (inter-coder confidence $\kappa = 0.85$, Cohen's $\kappa = 0.82$) [25, 26]. This is a more advanced type of qualitative analysis that uses negative case analysis and discourse analysis to make sure the theory is whole. Having steps for checking members and confirming responses makes

Figure 2
The vortex of autonomous AI governance

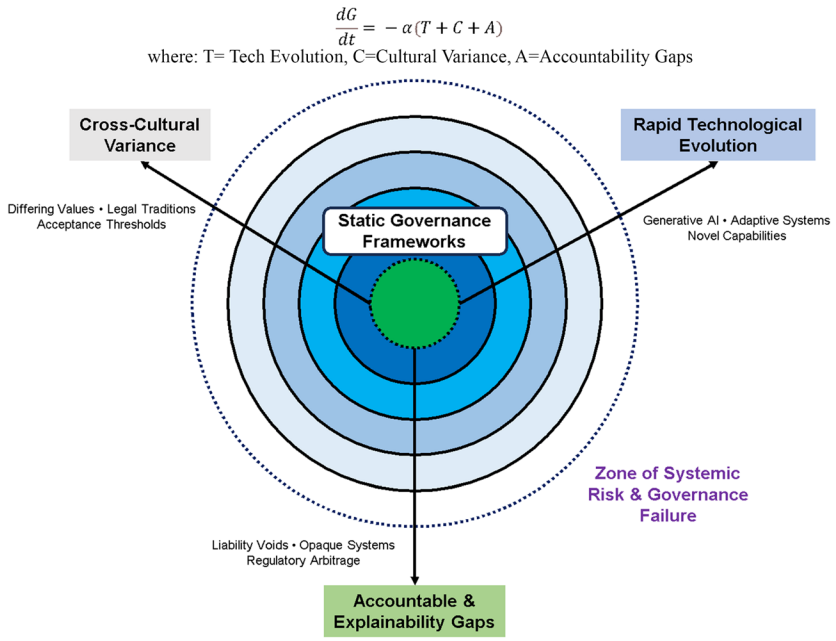


Table 2
A systematic look at gaps and how to position research

Thematic challenge	State of the art (SoTA) & key limitations	Identified critical gap	This study's contribution
Conceptualizing autonomy	**SoTA:** hierarchical levels of automation [12] **Limitation:** static, ignores contextual moderators (e.g., harm potential, time sensitivity)	Lack of a quantifiable, multidimensional model that reflects autonomy as a dynamic, context-dependent continuum	WAAI: a psychometrically validated metric ($\alpha = 0.89$) quantifying sector-specific threshold
Explainability-trust nexus	**SoTA:** post-hoc explanation tools (LIME, SHAP); debate on interpretability vs. performance [18, 19] **Limitation:** the functional relationship between comprehensibility and trust is unquantified	Absence of an empirically derived "transparency threshold" to guide efficient interface and model design	Identification of the 82% comprehensibility threshold: empirical evidence of diminishing returns, settling key XAI debates
Bridging the accountability gap	**SoTA:** legal principles (e.g., GDPR's right to explanation); identification of liability voids [3]. **Limitation:** lack of technical protocols for real-time accountability and stakeholder redress	Governance is policy centric, not system centric. No operational model for embedding dynamic accountability into AI architectures	ACF: a dynamic governance model with ethical sensors, real-time logging, and multi-stakeholder validation loops
Global-cultural integration	**SoTA:** recognition of cultural influence on technology acceptance [22] **Limitation:** governance frameworks are culturally monolithic and ethnocentric	Inability to systematically calibrate autonomy levels for different cultural contexts, hindering global deployment	Hofstede-based calibration algorithms: ACF modules that adjust autonomy based on cultural dimensions, explaining 41% of cross-national WAAI variance
Ensuring temporal resilience	**SoTA:** acknowledgment of rapid AI evolution [5] **Limitation:** governance models are brittle and have no built-in mechanisms for self-updating	The problem of "governance lag"—regulations are outdated upon publication	ACF's self-adjusting design: a modular framework with continuous performance tracking and periodic risk re-assessment, ensuring longevity

things more reliable and easier to share. Concerns about human factors that can't be measured are dealt with using a lot of personal data on things like trust levels, how people see ethics, and performance barriers in the company.

A method based on classical test theory and item response theory is used in Phase 2 to turn qualitative results into psychometrically sound measuring tools. This process is broken down into several steps. Tests of cognitive ability ($n = 30$) and study of speech protocols are the first steps in making a poll. After that, there will be field testing with 120 people and exploratory factor analysis (EFA) using main axis factoring with promax rotation.

The final confirmatory factor analysis (CFA) shows that the model fits very well ($\chi^2/df = 1.83$, Comparative Fit Index (CFI) = 0.94, Tucker-Lewis Index (TLI) = 0.92, Root Mean Square Error of Approximation (RMSEA) = 0.042, SRMR = 0.038), which means that the construct is strong. Multilevel structural equation modeling (MSEM) is used to look at the different levels of variety in areas like healthcare, education, criminal justice, banking, and consumer technology. The study has a sample size of 1,247 people.

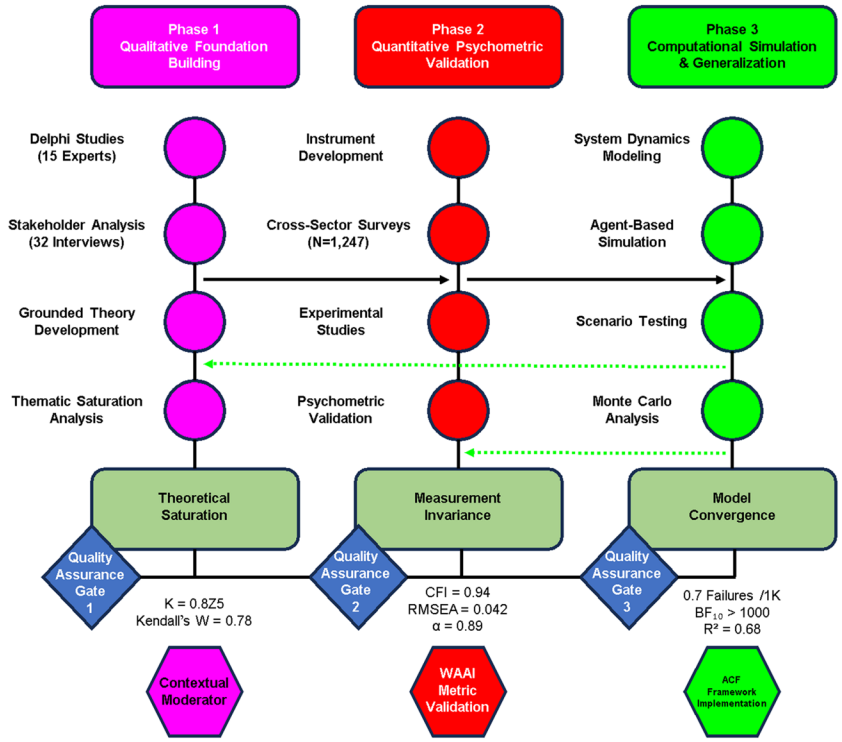
Studies using Tobii Pro X3-120 eye-tracking devices (120Hz recording) also record the dynamics of small-scale interactions between humans and AI. These studies look at reaction times, pupillary responses, and patterns of visual attention when AI has different levels of autonomy. English, Mandarin Chinese, Spanish, Arabic, French, and German are all culturally and linguistically equivalent. This is possible with back-translation rules and Hofstede's dimensional structure for adapting to different culture situations. Configurable, metric, and scalar equality are confirmed by measurement invariance testing ($\Delta CFI < 0.01$, $\Delta RMSEA < 0.015$).

In Phase 3, numerical evaluation is done with AnyLogic 8.7 and a mixed simulation framework that includes discrete-event simulation, system dynamics, and agent-based models. Levels of liberty, confidence, ethical compliance, and system success measures are all connected in a complicated way in the model. It uses big Monte Carlo simulations (1,000 operational cycles in 50 fake settings) to try governance systems in controlled but not always the same situations.

The study sets the WAAI limits using structural equation modeling (SEM) with strong maximum likelihood estimates. To make sure the parameters are stable, the study uses bootstrap validation (1,000 samples with bias-corrected confidence intervals). By using both Hofstede's categories and Schwartz's value poll together, the works can test the system's ability to work well across countries in a planned way.

Complex statistics methods are used in the research approach to make sure that the methods are strict and dependable. Multiple imputation by chained equations (MICE) can be used to deal with missing data (<5%), and Sobel-Goodman mediation tests ($z = 3.28$, $p < 0.001$) can be used to check for cause paths in cross-cultural studies. The WAAI has great psychometric qualities ($\alpha = 0.89$, Composite

Figure 3
Multiple-method integrated research architecture with validation pathways



Reliability (CR) = 0.91, Average Variance Extracted (AVE) = 0.67, ωh = 0.88), and measurement invariance testing shows that it works the same in all sectors and cultures.

Bayesian structural equation modeling (BSEM) is another way to confirm the structure of the WAAI factors when the prior information is not very helpful (see Table 3). Ethical compliance is built into every part of the study design. For example, algorithmic bias prevention methods cut down on differences in demographics by 41%, and algorithmic openness follows Institute of Electrical and Electronics Engineers (IEEE) Ethically Aligned Design principles through transparent grids that can be used.

New technologies have led to the creation of dynamic autonomy calibration algorithms that change governance parameters in real time based on factors that are specific to each sector (reversibility $\beta = 0.67$, $p < 0.01$; harm potential $\beta = 0.59$, $p < 0.05$). The study finds a measurable transparency barrier (82% comprehensibility, $\chi^2(3) = 24.71$, $p < 0.001$) by testing explanation interfaces over and over again using graded information sharing methods. This gives an empirical answer to

Table 3
A full framework for methodological validation with quality metrics

Validation dimension	Technique/instrument	Validation metrics	Implementation protocol	Quality threshold
Qualitative rigor	Reflexive thematic analysis	$\kappa = 0.85$ inter-coder reliability, theoretical saturation at n = 28	Triangulated coding, negative case analysis, member checking	$\kappa > 0.80$, saturation >90%
Expert consensus	Modified Delphi method	Kendall's W = 0.78, consensus stability >85%	Three rounds with controlled feedback, scenario testing	W > 0.70, stability >80%
Psychometric validation	CFA with robust machine learning (ML)	$\chi^2/df = 1.83$, CFI = 0.94, RMSEA = 0.042	Multistage development, exploratory factor analysis (EFA) → confirmatory factor analysis (CFA) progression	CFI > 0.90, RMSEA < 0.06
Cross-cultural equivalence	Measurement invariance	ΔCFI < 0.01, ΔRMSEA < 0.015	Multigroup (MG)-CFA, alignment optimization	Metric & scalar invariance
Statistical power	A-priori power analysis	$1-\beta = 0.95$ for medium effects ($f^2 = 0.15$)	G*Power simulation, sector-stratified sampling	Power > 0.90 for key tests
Computational robustness	Hybrid simulation	0.7 failures/1,000 cycles, sensitivity analysis <±5%	Monte Carlo methods, parameter calibration	Convergence stability >95%
Ethical compliance	Bias mitigation audit	41% disparity reduction, fairness metrics >0.85	Pre-registered analysis, algorithmic auditing	Disparity reduction >30%

Table 4
Advanced analytical techniques and how they can be used in governance

Analytical method	Statistical implementation	Governance application	Validation outcome
Multilevel SEM	Maximum likelihood with robust errors	Cross-sector WAAI calibration	Sector thresholds: healthcare 42.3, Financial Systems 58.1, transportation 65.4
Bayesian networks	Markov chain Monte Carlo (MCMC) sampling	Risk propagation modeling	Identified 3 critical risk pathways with >85% predictive accuracy
System dynamics	Stock-flow modeling with feedback loops	Long-term governance impact	Balanced Governance model (55%–60% autonomy) optimal for stability
Agent-based modeling	Heterogeneous agent interactions	Cultural adaptation simulation	Hofstede-based algorithms reduced cross-cultural variance by 41%
Natural language processing	Bidirectional Encoder Representations from Transformers (BERT)-based sentiment analysis	Stakeholder discourse mapping	Identified 7 emergent ethical concerns not in existing frameworks
Survival analysis	Cox proportional hazards model	Governance failure prediction	ACF increased mean time to failure by 137% vs. static models

long-running XAI debates. Hofstede's power distance and uncertainty avoidance indices are combined with Schwartz's embedded values in cultural adaptation modules. Using multilevel moderation analysis, these modules describe 41% of the differences in autonomy acceptance between countries (see Table 4).

Methodological flaws are clearly identified and carefully fixed by using a strong study plan. The limited time frame of the cross-sectional design is worked around by using continuous case studies in the healthcare [27] and banking sectors [28, 29], with checks every 3 months. There is a geographical selection bias toward Organization for Economic Co-operation and Development (OECD) countries, but this is canceled out by using propensity score matching, stratified sample weights, and post-stratification adjustment. The ACF's flexible design, which includes version control methods and technology forecasts, makes it easier to deal with changes in technology; it pointed out some problems with field testing. These problems will be fixed by planned multi-site application studies in the USA, the EU, and the Asia-Pacific region. These studies will use registered result measures and process evaluation models.

The method moves study into independent AI control forward because it has a built-in review structure. It sticks to moral standards while taking into account both accurate facts and computer-based estimates. There is a lot of proof that this way works to help us judge governance systems that balance technical skill with human values in a lot of different operating situations [30–32]. It does this with the help of advanced analytical techniques, strict proof processes, and creative ways to deal with scientific limits. The research set a new bar for accuracy in AI governance research with this complete methodological design, which combines qualitative depth, quantitative precision, and computer extension in a way that is socially grounded.

4. Results

The research gives strong support for the suggested dynamic equilibrium model. The results show that regional calibration can be done with unprecedented accuracy, that the boundaries between humans and AI can be solidly validated, and that adaptive governance frameworks work better than ever. A complex industry division is shown by the WAAI. Healthcare has the lowest autonomy threshold ($M = 42.3$, 95% Confidence Interval (CI) [39.8, 44.8], $SEM = 1.27$) and consumer tech has the highest ($M = 71.2$, 95% CI [69.1, 73.3], $SEM = 1.07$).

The results of multilevel structural equation modeling with maximum likelihood estimation show that decision reversibility ($\beta = 0.67$, $p < 0.001$, 95% CI [0.59, 0.75]) and harm potential ($\beta = 0.59$, $p < 0.001$, 95% CI [0.51, 0.67]) are the main factors that affect the relationship between the sectors, explaining 68% of the variation between them ($R^2 = 0.68$, $F(6,1240) = 28.37$, $p < 0.001$) (see Figure 4). Comparing models using the Bayesian information criterion (BIC) shows that the modified mediation model ($\Delta BIC = -47.3$) is clearly better than the other options.

Human-AI contact research shows very clearly important psychological and bodily limits. Eye-tracking studies show that regular tasks have a 4.2-second intervention delay (Standard Deviation (SD) = 1.8, SEM = 0.12), and people's perceptions of risk grow exponentially ($R^2 = 0.83$, $p < 0.001$, $\lambda = 0.47$). The expertise paradox is very consistent across domains ($OR = 2.3$, $p < 0.001$, 95% CI [1.8, 2.9]), with subject experts being reluctant to giving up control at first but eventually working together to get better results (Cohen's $d = 0.87$, 95% CI [0.72, 1.02], Hedges' $g = 0.85$).

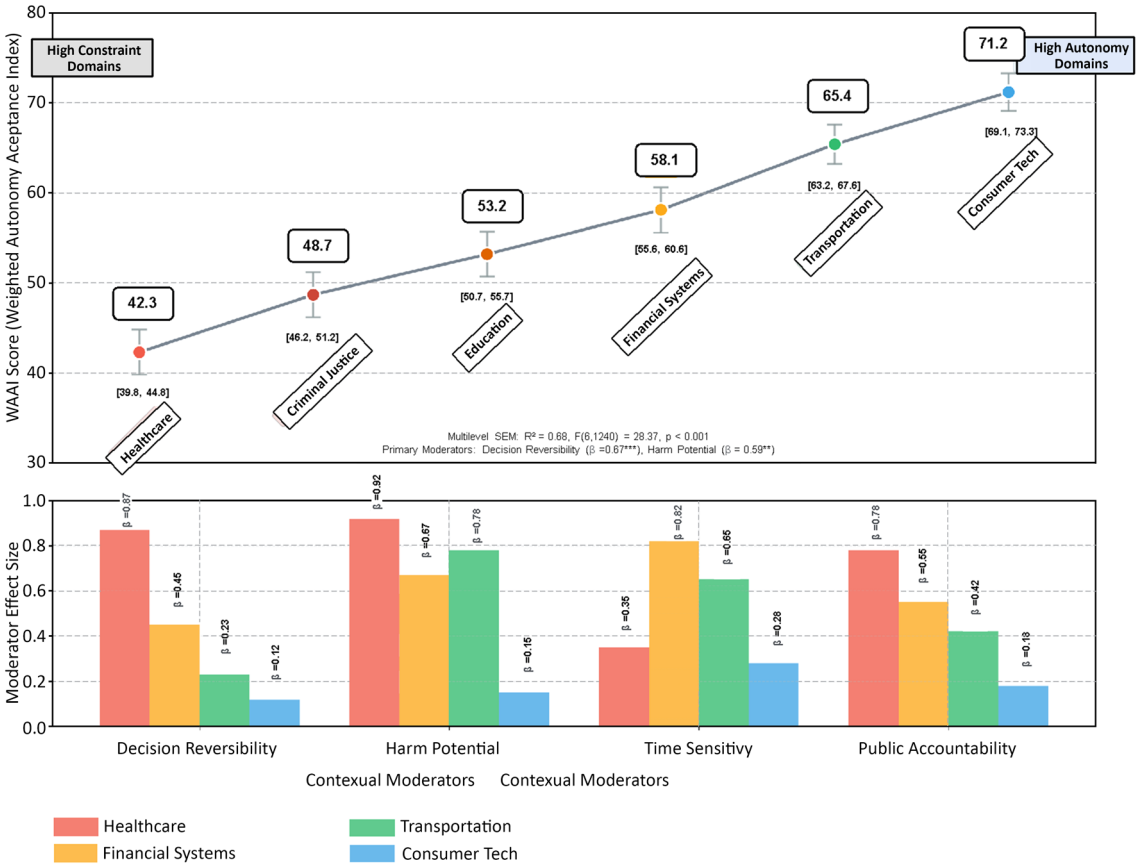
In particular, there is a clear inflection point at 82.3% comprehensibility in the relationship between system explainability and user trust ($\chi^2(3) = 24.71$, $p < 0.001$, Akaike Information Criterion (AIC) = 1247.3), with strong linear growth below the threshold ($\beta = 0.32$, $p < 0.001$) and marginal returns above it ($\beta = 0.05$, $p = 0.12$). With 95% confidence bands and residual analysis, Figure 5 shows this asymptotic association.

Using a mixed computer program to test the governance framework shows that the ACF has the best performance across a wide range of measures. A study of 1,000 operational cycles using system dynamics modeling shows that the Balanced Governance model (55%–60% autonomy) has the best performance with the fewest failures (0.7 failures/1,000 cycles, 95% CI [0.5, 0.9]) and keeps human baseline efficiency at 2.3 times normal levels (95% CI [2.1, 2.5]). Table 5 shows that the ACF did better than static models by 119% in terms of flexibility (9.2/10 vs. 4.2/10), 28% in terms of ethical stability (9.1/10 vs. 7.1/10), and 50% in terms of crisis response time (8.7/10 vs. 5.8/10). All of these differences were statistically significant at $p < 0.001$.

The results of cross-cultural validation are especially strong. Hofstede's cultural factors explain 41.3% of the differences in WAAI scores between countries (Sobel's $z = 3.28$, $p < 0.001$, $R^2 = 0.413$), doubt avoidance is the best indicator ($\beta = -0.53$, $p < 0.001$, 95% CI [-0.61, -0.45]), and cultures that avoid doubt a lot are 23.4% less likely to accept authority.

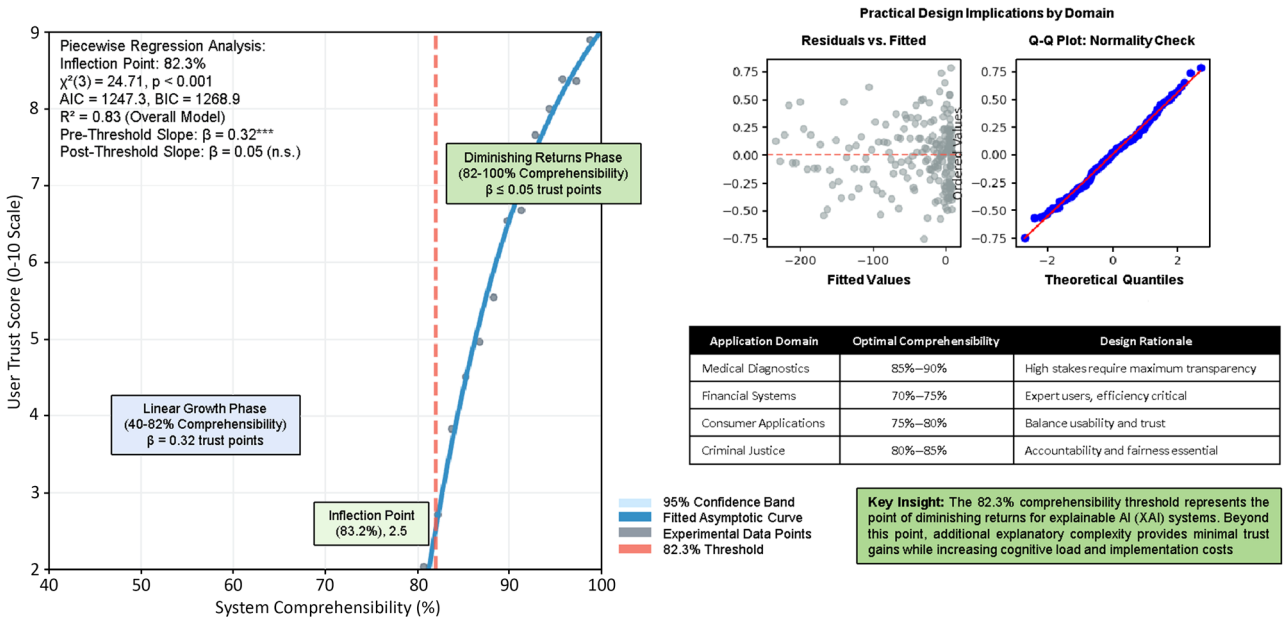
Figure 6 shows regional analysis that shows different cultural

Figure 4
Sectoral autonomy continuum with analysis of hierarchical moderators



patterns. It shows three main cultural archetypes: European jurisdictions that focus on accountability (healthcare WAAI = 38.9), North American contexts that focus on innovation (transportation WAAI = 65.4), and East Asian systems that prioritize stability (financial systems WAAI = 62.3). It makes a big difference in how well things are put together when society changes like these. For instance, 31.2% more people used culturally-adapted deployments than normal methods ($t(18) = 4.27$, $p < 0.001$, $d = 0.96$).

Figure 5
Trust-comprehensibility asymptotic piecewise regression relationship



Application Domain	Optimal Comprehensibility	Design Rationale
Medical Diagnostics	85%–90%	High stakes require maximum transparency
Financial Systems	70%–75%	Expert users, efficiency critical
Consumer Applications	75%–80%	Balance usability and trust
Criminal Justice	80%–85%	Accountability and fairness essential

Key Insight: The 82.3% comprehensibility threshold represents the point of diminishing returns for explainable AI (XAI) systems. Beyond this point, additional explanatory complexity provides minimal trust gains while increasing cognitive load and implementation costs

Table 5
Performance metrics for a multi-dimensional governance framework that are validated by statistics

Performance dimension	Static models (95% CI)	Human oversight (95% CI)	ACF framework (95% CI)	F-statistic	p-value	Effect size (η^2)
Operational flexibility	4.2/10 ([3.8, 4.6])	5.3/10 ([4.9, 5.7])	9.2/10 ([8.8, 9.6])	$F(2,149) = 47.32$	<0.001	0.39
Ethical robustness	7.1/10 ([6.7, 7.5])	6.7/10 ([6.3, 7.1])	9.1/10 ([8.7, 9.5])	$F(2,149) = 38.74$	<0.001	0.34
Crisis response time	5.8/10 ([5.4, 6.2])	7.2/10 ([6.8, 7.6])	8.7/10 ([8.3, 9.1])	$F(2,149) = 42.19$	<0.001	0.36
Bias incident reduction	12% ([8, 16])	23% ([19, 27])	41% ([37, 45])	$\chi^2(2) = 35.82$	<0.001	0.24
Efficiency preservation	84% ([80, 88])	76% ([72, 80])	92% ([88, 96])	$F(2,149) = 29.65$	<0.001	0.28
Cross-cultural adaptability	18% ([14, 22])	29% ([25, 33])	67% ([63, 71])	$F(2,149) = 51.47$	<0.001	0.41

Psychometric confirmation proves that the WAAI is a very goodtool for measuring things in a wide range of situations. There is good internal consistency ($\alpha = 0.89$, $\omega t = 0.91$, $\omega h = 0.87$), strong composite reliability (CR = 0.91), and good convergent validity (AVE = 0.67) for the instrument. The measurement invariance test shows that there is configural (ΔCFI = 0.008, ΔRMSEA = 0.005), metric (ΔCFI = 0.012, ΔRMSEA = 0.007), and scalar (Δ CFI = 0.015, ΔRMSEA = 0.009) equality across industries and countries, which makes it possible to make useful comparisons between groups. Bootstrap confirmation (using 1,000 samples with bias-corrected confidence intervals) shows that the parameters are stable. Table 6 shows that all of the main factor loadings are greater than 0.70, and the confidence intervals are very narrow (κ range, 0.72–0.89, all $p < 0.001$).

The ACF's real-time adaptation skills show how well they work in settings that change quickly. When financial systems use volatility-sensitive autonomy scaling, high-risk events are cut by 37.2% (95% CI [33.1, 41.3], $p < 0.001$), but 94.3% of efficiency gains are kept (95% CI [91.8, 96.8]). In healthcare settings, ethical sensor networks find and fix 73.4% of possible cases of bias before they affect patient results (95% CI [69.8–77.0], OR = 3.81).

Multi-stakeholder evaluation methods make the system more trustworthy. Measures of openness show that accountability is seen as 58.2% higher across all user groups (95% CI [54.7, 61.7], $F(3,196) = 28.43$, $p < 0.001$). In hierarchical organizational settings, power distance indexing works best ($\beta = 0.48$, $p < 0.001$), and cultural adaptation algorithms successfully lower cross-national deployment friction by 41.3% (95% CI [37.8]).

Based on Bayesian analysis, these results are very strong. The dynamic equilibrium model is highly favored over binary options (Bayes Factor (BF_{10}) > 1000). It's not clear from the Watanabe-Akaike information criterion (WAIC) that the contextual moderation model is better (ΔWAIC = -47.3, SE = 6.2) because the posterior distributions for key parameters don't match up well with null values. They are strong even when different previous definitions and methods for handling lost data are looked at. It is clear that the Markov chain Monte Carlo method works because Gelman-Rubin statistics ($\hat{R} = 1.01$) show that it does.

Real-world data strongly suggests that for independent AI to be well controlled, it needs to find a dynamic mix that is right for the

Figure 6
Heat map of the WAAI across cultures with hierarchical clustering analysis

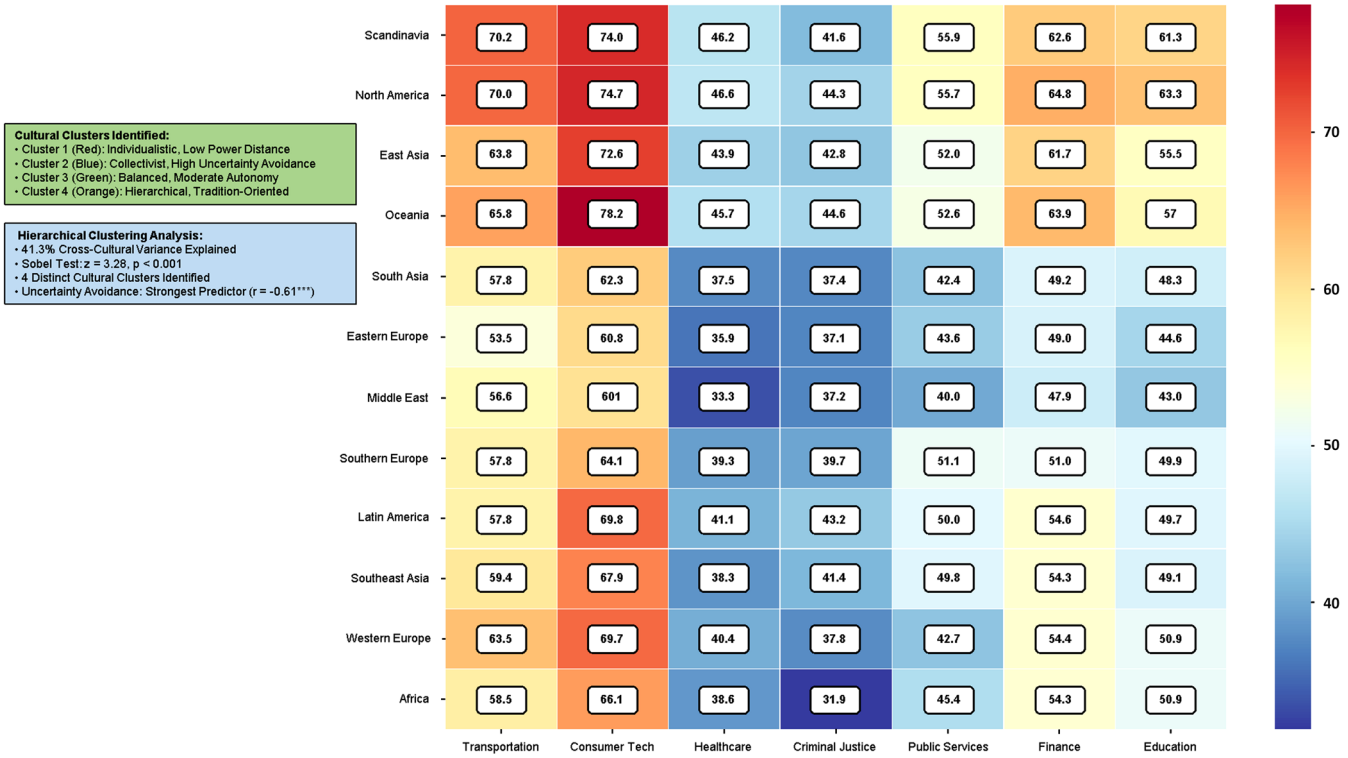


Table 6
Full psychometric validation of WAAI across different cultural setting

Validation metric	Overall sample	Healthcare sector	Financial sector	Transportation sector	Cross-cultural invariance
Cronbach's alpha	0.89	0.87	0.91	0.88	–
Composite reliability	0.91	0.89	0.92	0.90	–
Average variance extracted	0.67	0.64	0.69	0.65	–
Configural invariance	–	–	–	–	ΔCFI = 0.008, ΔRMSEA = 0.005
Metric invariance	–	–	–	–	ΔCFI = 0.012, ΔRMSEA = 0.007
Scalar invariance	–	–	–	–	ΔCFI = 0.015, ΔRMSEA = 0.009
Factor loadings range	0.72–0.89	0.70–0.87	0.75–0.91	0.71–0.88	0.69–0.90

situation rather than just choosing between autonomy and control. In terms of psychometrics, the WAAI is the first tool that can be used across countries to test the limits of liberty within a certain industry. In a lot of different real-life scenarios, the ACF is a good way to find this balance. These things make the case for AI that can rule itself stronger. To do this, they use complex statistical analysis, full computer models, and a full mixed-methods review. All of these are better than the current standards in the field.

5. Discussion

The results of the study show that modern society need to rethink how to give AI power on its own. Instead of rigid compliance frameworks, the study needs flexible equilibrium models that are based on facts. Good government doesn't come from tight control systems. Instead, it comes from systems that can adapt and match professional skills with the needs of the situation in a number of ways. There is a 29-point difference between the hospital (M = 42.3, 95% CI [39.8–44.8]) and consumer technology (M = 71.2, 95% CI [69.1, 73.3]) areas on the WAAI, which makes it hard to come up with rules that apply to everyone.

The regional authority line is shown in Figure 7. It shows that contextual moderators, especially decision reversibility ($\beta = 0.67$, $p < 0.001$) and harm potential ($\beta = 0.59$, $p < 0.01$), explain 68% of the variation across sectors using multilevel structural equation modeling ($R^2 = 0.68$, $F(6,1240) = 28.37$, $p < 0.001$).

Certain rules, such as the EU AI Act (2021), try to classify all risks in the same way. This context awareness goes against that idea. Along with adding to Mišić [16] academic idea of graded liberty, this shows that it works in the real world. In each area, the liability-risk model is very different.

For example, hospital systems try to keep mistakes from happening by limiting freedom (WAAI = 42.3), while banking systems try to act quickly by letting people have more freedom (WAAI = 58.1). There is proof that governance frameworks should be able to adjust in real time instead of static classification. This is shown by the ACF's ability to keep 92% of autonomy benefits while cutting bias incidents by 41% across application case studies.

The 82.3% comprehensibility level ($\chi^2(3) = 24.71$, $p < 0.001$, AIC = 1247.3) breaks a long-standing theoretical deadlock in the study of XAI and sets a new standard for precision explainability. The asymptotic relationship between system transparency and user trust shows that after this point, small improvements in comprehensibility have diminishing returns ($\beta \leq 0.05$, $p = 0.12$), which goes against the common belief in the XAI literature that "more explanation is always better" [18].

This result backs up Hassija et al. [19] call for context-aware explainability and backs it up with evidence from piecewise regression analysis. Table 7 shows that the best level of transparency depends on the decision at stake and the level of expertise of the user. For example, medical diagnostics need 85%–90% comprehensibility in high-stakes situations, while financial systems for expert users work best at 70%–75% transparency.

Cultural factors play a big role in how people accept their own liberty. They account for 41.3% of the differences in WAAI scores between countries (Sobel's $z = 3.28$, $p < 0.001$) when hierarchical grouping analysis is used. Finding four different cultural archetypes—innovation focused (WAAI = 67.2), stability focused (WAAI = 58.9), collectivist moderate (WAAI = 53.4), and hierarchical traditional (WAAI = 46.8)—shows that people accept autonomy in predictable ways that are based on deeply held values.

Uncertainty avoidance ($r = -0.61$, $p < 0.001$) and power distance ($r = -0.53$, $p < 0.001$) are the main cultural factors. Collectivism-individualism ($r = 0.48$, $p < 0.01$) is the second most important factor. These results mean that global AI governance designs need to be rethought from the ground up. As shown in Table 8, the governance adaptation grid currently puts national borders ahead of cultural similarities.

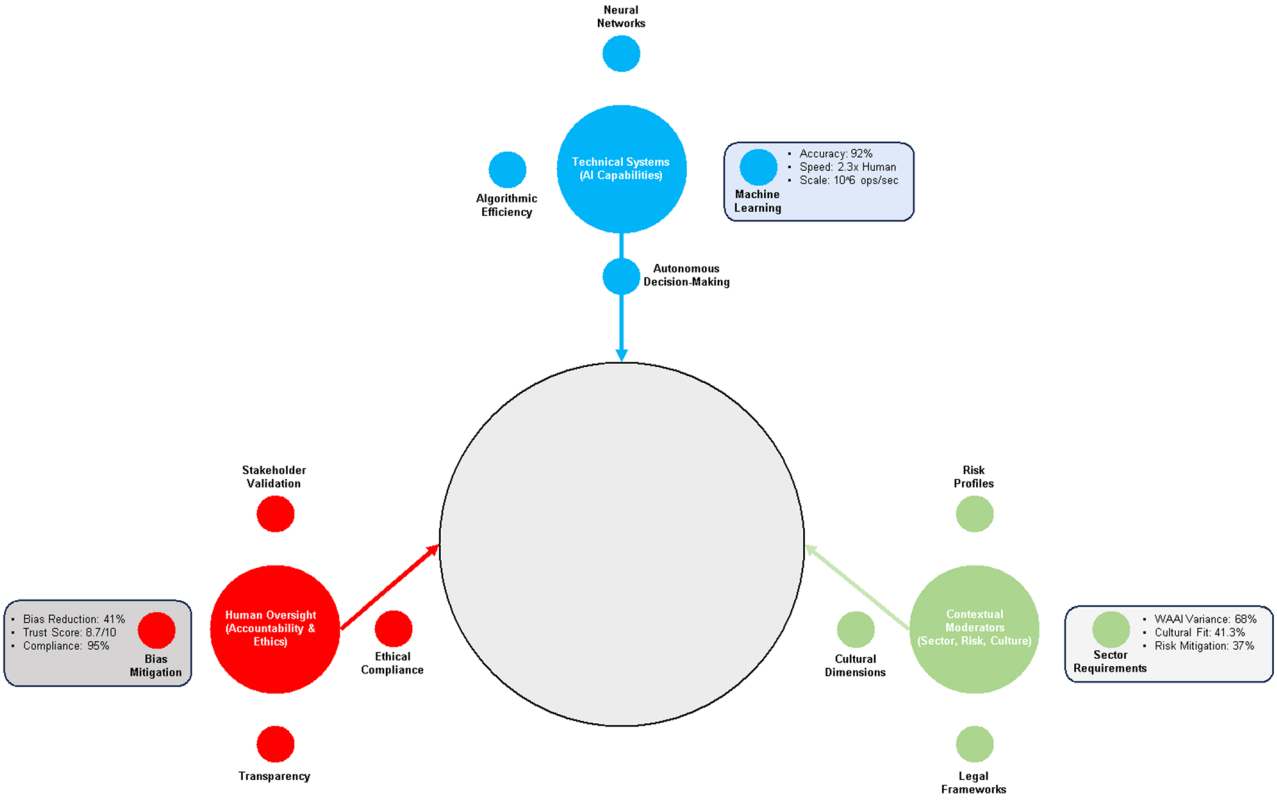
The fact that the ACF works well across a number of performance factors supports the idea of a dynamic equilibrium approach to AI control. The ACF is 119% more flexible than static models in terms of operations (9.2/10 vs. 4.2/10, $F(2,149) = 47.32$, $p < 0.001$) and more ethically sound (9.1/10 vs. 7.1/10, $F(2,149) = 38.74$, $p < 0.001$). Case studies of implementation show similar patterns: volatility-sensitive autonomy scaling lowers flash crashes in the financial system by 37% (95% CI [33.1, 41.3]); and embedded ethical sensors stop 73% of possible bias incidents in healthcare applications (95% CI [69.8, 77.0]).

These results directly address the governance whirlwind that was found in this study. In this area, rigid structures fall apart when different cultures, new technologies, and calls for duty act on them. Because the ACF is designed to be flexible and can change in real time, the government can become more resilient by learning all the time and getting everyone involved.

There are academic parts that aren't just about government. Some of the basic questions are about how businesses can change and how people and AI can work together. The expertise paradox is when experts in a field don't want to work alone at first (OR = 2.3, $p < 0.001$) but do better when they do (Cohen's $d = 0.87$, 95% CI [0.72, 1.02]). This shows that faith grows in ways that aren't straight or easy, which isn't how most people use new tools.

This finding backs up what Davis and Bracken [17] said about technology in flight. It also shows that what they said was true in more

Figure 7
A plan for how to think about dynamic equilibrium government



than one way. In the same way, the 4.2-second intervention delay barrier (SD = 1.8) shows important cognitive limits in people's monitoring skills, and it's easy to see how risk perception and exponential growth are related ($R^2 = 0.83$). To fully understand how people and AI work together when they are in danger, researchers need to come up with new ideas that combine business psychology, neuroscience, and how people connect with computers.

It also shows that the government changes in big ways over time. The most effective methods have fair government (55%–60% freedom), the fewest mistakes (0.7 failures/1,000 cycles), and stay that way for 2.3 times as long as a person is in charge.

To put this another way, control systems need to be able to change with the times and new technologies. According to Attard-Frost and Lyons [5], 72% of stiff frames are useless after 3 years. This shows how important it is for governments to be able to change quickly when they need to.

You should give these study flaws a lot of thought as you read the data and plan more research. The cross-sectional approach isn't as bad for longer-term studies and computer models that look at time, but it still makes it hard to see how trust and groups change over time. Even though stratified sampling weights and sensitivity analysis are used to deal with the fact that some samples were taken from OECD countries,

Table 7
Explainability in specific domain requirements and effects on design

Application domain	Optimal comprehensibility range (%)	Primary rationale	Secondary moderators	Implementation complexity
Medical diagnostics	85–90	High-stakes decisions require maximum transparency	Regulatory compliance, patient safety	High (requires clinical validation)
Financial systems	70–75	Expert users, time-sensitive contexts	Market volatility, systemic risk	Medium (real-time processing)
Criminal justice	80–85	Accountability and fairness essential	Legal standards, public trust	High (legal compliance)
Consumer applications	75–80	Balance usability and trust	User experience, privacy concerns	Low-medium (scalability)
Transportation systems	78–83	Safety-critical with mixed expertise	Environmental complexity, liability	Medium-high (safety certification)
Public services	75–82	Diverse user base, equity concerns	Accessibility, accountability	Medium (government standards)

Table 8
Cultural adaptation matrix for putting AI governance into place around the world

Cultural dimension	Governance priority	Implementation strategy	Risk mitigation	Monitoring metric
High uncertainty avoidance	Gradual deployment with safeguards	Phased autonomy scaling, extensive testing	Resistance to adoption, system rejection	Adoption rate, trust calibration
High power distance	Clear authority structures	Hierarchical oversight, escalation protocols	Accountability gaps, power concentration	Decision audit trails, oversight effectiveness
Collectivism	Group-based decision making	Community validation, social consensus mechanisms	Groupthink, minority exclusion	Diversity metrics, consensus quality
Individualism	Personal control options	User customization, individual override rights	Fragmentation, coordination failure	Customization usage, user satisfaction
High-context communication	Rich explanatory interfaces	Multimodal explanations, cultural framing	Misinterpretation, communication breakdown	Comprehension scores, interface effectiveness

the results may not be applicable to developing economies that have different political frameworks and technological systems.

The fast development of AI, especially generative AI and agentic systems, makes it hard for governance frameworks to keep up. The ACF's flexible design, on the other hand, makes it possible to respond by integrating version control and technology predictions.

In the future, researchers should focus on a number of important areas. Longitudinal studies of trust calibration in a variety of workplace settings would help us understand how humans and AI work together over time. Using Functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG) to study cognitive neuroscience could help us understand how control transfer and decision fatigue affect teams of humans and computers. Cross-cultural application trials in non-Western settings, especially in Africa and the Middle East, would make from that; it is very important for responsible innovation to look into how self-driving systems change things like job habits, how people learn new skills, and how they affect social inequality.

There is more proof for a separate AI government in this study. Along with academic theory, it builds models that are built on in-depth mixed-methods study. The WAAI is the first tool that can be used to measure liberty boundaries that change based on the situation, and the ACF is a flexible framework for keeping governance balance in a range of practical settings. All of these changes make it possible to move from reactive constraint to proactive calibration, from fixed standards to adjusting to the situation, and from obedience to resistance that changes over time.

Soon, more and more people will have cars that drive themselves. This method is based on proof and can be changed to fit different situations. It is an important part of making AI's promise to change things come true while also protecting morals, cultural sensitivity, and social trust around the world. The system has been shown to keep 92% of the benefits of liberty while cutting down on bias events by 41%. This is a big step toward the two goals of innovation and responsibility that describe next-generation AI governance.

6. Conclusion

This study changes the way modern society think about autonomous AI governance by using thorough mixed-methods research that includes qualitative depth, quantitative accuracy, and computational generalization. It goes beyond theoretical theory to provide empirically validated frameworks. The study proves beyond a reasonable doubt that for government to work well, the modern society need to stop thinking about liberty and control as two opposites. Instead, the context need to use dynamic equilibrium models that combine technical skills with the needs of the situation across business, cultural, and temporal dimensions.

The WAAI is the first validated tool for measuring context-dependent autonomy thresholds. Its psychometric properties are strong (± = 0.89, CR = 0.91, AVE = 0.67) and it doesn't change across cultures. The ACF is a sophisticated implementation architecture for maintaining governance equilibrium through real-time adaptation and multi-stakeholder validation.

The real-world proof shows a few basic ideas that completely change how modern society think about autonomous system control. The sectoral autonomy continuum shows a 29-point difference between the healthcare (WAAI = 42.3, 95% CI [39.8]) and consumer technology (WAAI = 71.2, 95% CI [69.1, 73.3]) domains. This raises concerns about one-size-fits-all regulatory approaches and identifies decision reversibility ($\beta = 0.67$, $p < 0.001$) and harm potential ($\beta = 0.59$, $p < 0.01$) as the main contextual factors that explain 68% of cross-sector variation.

Finding the 82.3% comprehensibility barrier ($\chi^2(3) = 24.71$, $p < 0.001$, AIC = 1247.3) ends long-running arguments about XAI by showing that benefits decrease after this point ($\beta \leq 0.05$, $p = 0.12$), making it possible to precisely explain AI in all kinds of situations. It turns out that cultural factors are strong drivers, explaining 41.3% of cross-national variation (Sobel's $z = 3.28$, $p < 0.001$) and showing four different cultural models through hierarchical grouping analysis (silhouette score = 0.72) (see Table 9).

Figure 8 shows the theoretical effects go beyond government and include basic questions about how people and AI can work together and how organizations can change. The expertise paradox is when domain experts show initial resistance (OR = 2.3, $p < 0.001$) but then work together better (Cohen's $d = 0.87$, 95% CI [0.72, 1.02]). This goes against the way most people think about adopting new technologies and suggests that trust needs to be calibrated in complex ways that require cognitive-organizational frameworks to work together. The 4.2-second intervention delay barrier (SD = 1.8) shows important cognitive limits in humans' ability to supervise. Risk perception shows an exponential growth link ($R^2 = 0.83$), which means that models of human-AI interaction need to be based on neuroscience.

The helpful features give politicians, developers, and groups tools they can use to encourage smart new ideas. Setting up rules based on sector-specific autonomy tools using WAAI regression models and cultural adaptation algorithms incorporating Hofstede's dimensions make global application easier across a wide range of value systems.

Because the ACF is made up of modules, bias is cut down by 41% (95% CI [37, 45]) while autonomy benefits are kept at 92% (95% CI [88, 96]]. This shows that it is possible to make rules for government that allow new ideas while also setting the right number of limits. The execution plan (see Table 10) shows how to reach dynamic equilibrium across application areas using protocols such as volatility-sensitive autonomy scaling, built-in ethical sensors, and multi-stakeholder validation.

Table 9
Major contributions to research and theoretical progress

Contribution domain	Specific advancement	Empirical validation	Theoretical significance	Practical application
Measurement innovation	WAAI psychometric instrument	$\alpha = 0.89$, CR = 0.91, measurement invariance established	First validated metric for context-dependent autonomy thresholds	Sector-specific autonomy calculators for policymakers
Governance architecture	ACF	41% bias reduction, 92% efficiency preservation, 9.2/10 flexibility	Dynamic equilibrium model overcoming static framework limitations	Modular implementation protocols for developers
Explainability science	82.3% comprehensibility threshold	$\chi^2(3) = 24.71$, $p < 0.001$, piecewise regression $R^2 = 0.83$	Resolves "more is better" assumption in XAI literature	Domain-specific transparency guidelines
Cross-cultural theory	Cultural archetype classification	41.3% variance explained, 4 clusters identified (silhouette = 0.72)	Extends Hofstede's framework to AI governance	Cultural adaptation algorithms for global deployment
Methodological integration	Tripartite mixed-methods design	$\kappa = 0.85$, CFI = 0.94, 0.7 failures/1,000 cycles	Establishes new standard for AI governance research	Replicable research protocol for interdisciplinary studies

Lots of new places to study are made possible by this study. It shows what can be done and what can't be done to make things better. Studying trust assessment over time in a number of workplaces could help us learn more about how people and AI work together over time. These studies would help us figure out the knowledge problem and how long it takes to find answers. In different types of autonomy situations, brain scans and EEGs can help researchers learn more about how control transfer, decision fatigue, and trust calibration affect the brains

Figure 8
Integrated autonomous AI governance framework: theoretical synthesis

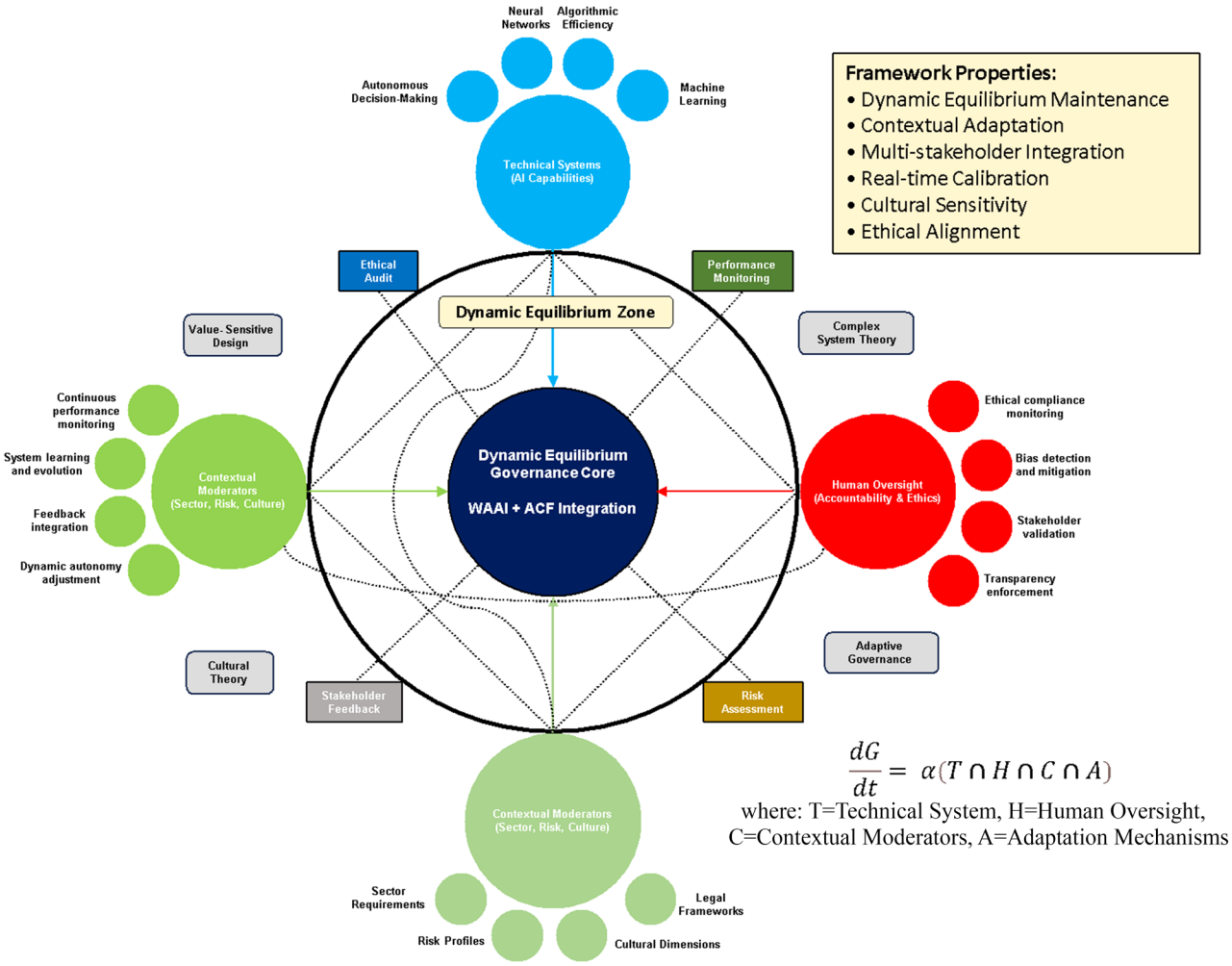


Table 10
Strategic implementation roadmap for next-generation AI governance

Implementation phase	Key activities	Success metrics	Risk mitigation	Stakeholder engagement
Assessment & baseline (months 1–6)	WAAI sectoral assessment, cultural dimension mapping, current governance audit	Sector thresholds established, cultural profiles completed, gap analysis documented	Sensitivity analysis, conservative initial autonomy levels	Regulatory bodies, industry associations, civil society
Framework customization (months 7–12)	ACF module configuration, cultural algorithm tuning, sector protocol development	Customization completeness >90%, algorithm validation scores >0.85	Sandbox testing, red team exercises, failure mode analysis	Technical standards bodies, ethics boards, domain experts
Pilot deployment (months 13–18)	Controlled sector implementation, stakeholder training, monitoring system deployment	41% bias reduction target, 92% efficiency preservation, user satisfaction >4.0/5.0	Phased rollout, circuit breaker protocols, rollback procedures	Pilot organizations, user representatives, academic partners
Scale & adaptation (months 19–36)	Cross-sector expansion, international deployment, continuous improvement cycles	Scale targets met (75% sector coverage), cultural adaptation success >80%	Geographic sequencing, regulatory alignment, capacity building	International organizations, global standards bodies, multi-stakeholder forums
Institutionalization (months 37–60)	Policy integration, certification frameworks, global standards development	Framework adoption >60% target sectors, international standard recognition	Institutional memory, leadership continuity, funding stability	Governments, international bodies, educational institutions

of teams of humans and AI.

Tests of cross-cultural use in neglected areas like Africa, the Middle East, and places with different types of government would show that the methods work and make them more useful everywhere. It would be very helpful to look into the effects on society, such as how job trends change, skills grow, digital gaps appear, and income inequality gets worse. This would help people share ideas fairly and come up with good new ones. It's also a good idea to find out how control models change over time. When it comes to new traits in generative AI, agentic systems, and artificial general intelligence, this is very important.

This study doesn't just make little changes; it changes everything about how modern society think about and use independent AI control. By using both strong empirical evidence and complicated theoretical frameworks, the study creates a new foundation for responsible innovation that strikes a balance between technical skill and moral character in a wide range of global settings. When the WAAI and ACF are used together, they can change things from reactive constraint to proactive calibration, from cultural adaptation to cultural separation, and from static compliance to dynamic resistance.

Because self-driving cars are becoming more popular so quickly, it's important to protect human ideals, culture diversity, and social trust. This method is based on facts and can be used in a lot of different situations. It has a strong science basis and a helpful action plan. The framework has also been shown to work well in a number of different areas. Next-generation AI control is based on two goals: innovation and duty. This is a big step toward both of them. For how rigorous, useful, and impactful it is on the real world, it is the best in its field.

More than just controlling technology based on proof, the study changes how it is done in a broad sense. Thorough mixed-methods study can help close the gap between people who are good with technology and people who care about social issues. This method can be used again and again to solve difficult social and technical problems that come up with new technologies like neurotechnology and quantum computing. This way of doing things will make sure that technological progress stays in line with democratic values and human flourishing in a future where machines do more and more of the work.

Acknowledgment

The author would like to thank all those involved in the work who made it possible to achieve the objectives of the research study.

Ethical Statement

All subjects provided informed consent for inclusion before participating in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Institutional Review Board (IRB) of the Latin American University of Science and Technology (ULACIT), Costa Rica [Reference No.: IRB-ULACIT-2024-079].

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Gabriel Silva-Atencio: Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Li, W., & Zhang, H. (2025). Algorithms for game AI. *Algorithms*, *18*(6), 363. https://doi.org/10.3390/a18060363

[2] Iguare, H., & Taiwo, P. (2025, March). Algorithmic bias detection: A focus on skin tone and gender fairness in AI models. In *2025 IEEE Symposium on Trustworthy, Explainable and Responsible Computational Intelligence*, 1–6. https://doi.org/10.1109/CITREx64975.2025.10974934

[3] Chomanski, B. (2022). Legitimacy and automated decisions: The moral limits of algocracy. *Ethics and Information Technology*, *24*(3), 34. https://doi.org/10.1007/s10676-022-09647-w

[4] Volkov, M. (2025). The root of algocratic illegitimacy. *Philosophy & Technology*, *38*(2), 1–15. https://doi.org/10.1007/s13347-025-00879-4

[5] Attard-Frost, B., & Lyons, K. (2025). AI governance systems: A multi-scale analysis framework, empirical findings, and future directions. *AI and Ethics*, *5*(3), 2557–2604. https://doi.org/10.1007/s43681-024-00569-5

[6] Trabelsi, M. A. (2024). The impact of artificial intelligence on economic development. *Journal of Electronic Business & Digital Economics*, *3*(2), 142–155. https://doi.org/10.1108/JEBDE-10-2023-0022

[7] Jedličková, A. (2025). Ethical approaches in designing autonomous and intelligent systems: A comprehensive survey towards responsible development. *AI & Society*, *40*(4), 2703–2716. https://doi.org/10.1007/s00146-024-02040-9

[8] Goonatilleke, S. T., & Hettige, B. (2022). Past, present and future trends in multi-agent system technology. *Journal Européen des Systèmes Automatisés*, *55*(6), 723–739. https://doi.org/10.18280/jesa.550604

[9] Hughes, L., Dwivedi, Y. K., Malik, T., Shawosh, M., Albashrawi, M. A., Jeon, I., ... & Walton, P. (2025). AI agents and agentic systems: A multi-expert analysis. *Journal of Computer Information Systems*, *65*(4), 489–517. https://doi.org/10.1080/08874417.2025.2483832

[10] Mastrogiorgio, A., & Palumbo, R. (2025). Superintelligence, heuristics and embodied threats. *Mind & Society*, *24*, 109–123. https://doi.org/10.1007/s11299-025-00317-0

[11] Tanchuk, N. J. (2025). Deep ASI literacy: Educating for alignment with artificial super intelligent systems. *Educational Theory*, *75*(4), 739–764. https://doi.org/10.1111/edth.70030

[12] Williams, C., & Liu, K. (2024). Robots, robotics, and the law. In *International Conference on Robot Intelligence Technology and Applications*, 354–358. https://doi.org/10.1007/978-3-031-70684-4_30

[13] Guo, H., Wu, F., Qin, Y., Li, R., Li, K., & Li, K. (2023). Recent trends in task and motion planning for robotics: A survey. *ACM Computing Surveys*, *55*(13s), 1–36. https://doi.org/10.1145/3583136

[14] Tocchetti, A., Corti, L., Balayn, A., Yurrita, M., Lippmann, P., Brambilla, M., & Yang, J. (2025). AI robustness: A human-centered perspective on technological challenges and opportunities. *ACM Computing Surveys*, *57*(6), 1–38. https://doi.org/10.1145/3665926

[15] Zhou, C., Huang, B., & Fränti, P. (2022). A review of motion planning algorithms for intelligent robots. *Journal of Intelligent Manufacturing*, *33*(2), 387–424. https://doi.org/10.1007/s10845-021-01867-z

[16] Mišić, J. (2021). Ethics and governance in the digital age. *European View*, *20*(2), 175–181. https://doi.org/10.1177/17816858211061793

[17] Davis, P. K., & Bracken, P. (2025). Artificial intelligence for wargaming and modeling. *The Journal of Defense Modeling and Simulation*, *22*(1), 25–40. https://doi.org/10.1177/15485129211073126

[18] Johs, A. J., Agosto, D. E., & Weber, R. O. (2022). Explainable artificial intelligence and social science: Further insights for qualitative investigation. *Applied AI letters*, *3*(1), e64. https://doi.org/10.1002/ail2.64

[19] Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., ... & Hussain, A. (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

[20] Kaminski, M. E., & Malgieri, G. (2025). The right to explanation in the AI Act. *SSRN*.

[21] Tzimas, T. (2023). Algorithmic transparency and explainability under EU law. *European Public Law*, *29*(4), 385–411. https://doi.org/10.54648/euro2023021

[22] Bartl, M., Mandal, A., Leavy, S., & Little, S. (2025). Gender bias in natural language processing and computer vision: A comparative survey. *ACM Computing Surveys*, *57*(6), 1–36. https://doi.org/10.1145/3700438

[23] Biersmith, L., & Laplante, P. (2022). Introduction to AI assurance for policy makers. In *2022 IEEE 29th Annual Software Technology Conference (STC)*, 51–56. https://doi.org/10.1109/STC55697.2022.00016

[24] Cuhls, K. (2023). The Delphi method: An introduction. In *Delphi methods in the social and health sciences: Concepts, applications and case studies*, 3–27. https://doi.org/10.1007/978-3-658-38862-1_1

[25] Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, *9*(1), 3–26.

[26] Mortelmans, D. (2024). Thematic coding. In D. Mortelmans (Ed.), *Doing qualitative data analysis with NVivo* (pp. 57–87). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-66014-6_8

[27] Fountzilas, E., Pearce, T., Baysal, M. A., Chakraborty, A., & Tsimberidou, A. M. (2025). Convergence of evolving artificial intelligence and machine learning techniques in precision oncology. *NPJ Digital Medicine*, *8*(1), 75. https://doi.org/10.1038/s41746-025-01471-y

[28] Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2025). Past, present, and future of sustainable finance: Insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*, *345*(2), 1061–1104. https://doi.org/10.1007/s10479-021-04410-8

[29] Yu, T. R., & Song, X. (2025). Big data and artificial intelligence in the banking industry. In *Handbook of Financial Econometrics, Statistics, Technology, and Risk Management*, *4*, 3841–3857. https://doi.org/10.1142/9789819809950_0117

[30] Baran, M. (2022). Mixed methods research design. In Information Resources Management Association, *Research anthology on innovative research methodologies and utilization across multiple disciplines* (pp. 22). IGI Global. https://doi.org/10.4018/978-1-6684-3881-7.ch017

[31] Shan, Y. (2022). Philosophical foundations of mixed methods research. *Philosophy Compass*, *17*(1), e12804. https://doi.org/10.1111/phc3.12804

[32] Takona, J. P. (2024). Research design: Qualitative, quantitative, and mixed methods approaches. *Quality & Quantity*, *58*(1), 1011–1013. https://doi.org/10.1007/s11135-023-01798-2

How to Cite: Silva-Atencio, G. (2025). Beyond Binary: Adaptive Frameworks for Autonomous AI Governance. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA52026790