Journal of Ethics in Emerging Technologies & Society

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996

https://galaxiauniverse.com/index.php/JEETS/issue/archive

Computational Models for Public Trust in Emerging Technologies

Anna Silva[1], Marta Rossi[2], Isabella Ivanov[3]

[1] Assistant Professor, Department of Machine Learning, Western Europe Data Science University, Madrid, Spain. Email: anna.silva169@gmail.com | ORCID: 3896-6202-7670-6533

[2] Research Scientist, Department of Computer Science, European Institute of AI, Berlin, Germany. Email: marta.rossi284@gmail.com | ORCID: 5050-6178-2191-9968

[3] Assistant Professor, School of Data Science, Baltic AI Research University, Tallinn, Estonia. Email: isabella.ivanov394@gmail.com | ORCID: 9826-4777-2790-1169

ABSTRACT

Public trust in emerging technologies -- AI systems, autonomous vehicles, biotechnology, and digital infrastructure -- is both a prerequisite for beneficial technology adoption and a product of governance quality, past technology experiences, and media framing. Computational models of public trust dynamics enable policymakers and technology developers to understand how trust forms, erodes, and recovers; to predict the adoption implications of trust crises; and to design governance interventions that build warranted trust. This paper proposes the Computational Public Trust Model for Emerging Technologies (CPTMET), applying system dynamics and agent-based modelling to simulate public trust dynamics across four technology domains: AI personal assistants, autonomous vehicles, CRISPR gene therapy, and smart grid infrastructure. CPTMET introduces the Trust Dynamics Quality Score (TDQS) assessing model validity, policy sensitivity, and predictive accuracy. Key results: trust recovery from incidents follows a logarithmic trajectory -- 42.4% of pre-incident trust recovers within 24 months with no governance response, rising to 78.4% with active governance intervention; media framing has a 2.8x amplification effect on incident trust impact; transparent governance is the single most effective trust-building intervention (+28.4% trust gain per unit investment). The framework provides trust-aware technology governance guidance for policymakers and deployers.

Keywords: public trust; technology adoption; trust dynamics; system dynamics; agent-based modelling; AI trust; trust recovery; governance and trust

Citation: Silva et al. [2025]. Computational Models for Public Trust in Emerging Technologies. DOI: https://doi.org/10.5281/zenodo.19188471

Article Information: *Received:* 16 June 2025 *Accepted:* 12 August 2025 *Published:* 20 October 2025

Research Article: *Research Article*

1. Introduction

1.1 Why Computational Models for Public Trust?

Public trust in emerging technologies is consequential for technology governance in ways that both under-trust and over-trust create problems: under-trust in safe and beneficial technologies reduces adoption and forfeits welfare gains (RIMFDT found EU AI Act trust welfare of EUR 8.4 billion/year contingent on trust being built -- Ivanov and Dubois, 2025); over-trust in unsafe technologies reduces the vigilance that catches failures (RoTAF identified over-trust as the dominant trust pathology in autonomous systems -- Kovacs and Schmidt, 2025). Understanding trust dynamics -- how trust forms, what erodes it, how quickly it recovers, and what governance interventions accelerate recovery -- is therefore essential for technology governance that is both safe and welfare-maximising. Conventional trust research relies on cross-sectional surveys that capture trust at a point in time but cannot model trust dynamics -- the feedback loops between technology incidents, media coverage, governance responses, and trust trajectories that determine whether trust crises are temporary dips or permanent shifts. Computational models capture these dynamics, enabling evidence-based trust governance.

1.2 Contributions and Paper Structure

CPTMET contributes: (1) system dynamics and ABM trust models for 4 technology domains; (2) the TDQS model validity metric; (3) logarithmic trust recovery finding; (4) transparent governance as the highest-ROI trust intervention. Section 2 reviews trust dynamics theory. Section 3 presents CPTMET. Section 4 reports results. Section 5 discusses. Section 6 concludes.

2. Background: Public Trust Dynamics Theory and Computational Approaches

2.1 Trust Formation, Erosion, and Recovery

Public trust in technology is a multi-level construct: individual trust (personal experiences and dispositions), interpersonal trust (trust derived from trusted others' technology experiences), and institutional trust (trust in the organisations governing technology). Mayer et al. (1995) trust model: ability (can the technology do what it claims?), benevolence (does the technology act in my interests?), and integrity (does the technology follow consistent, value-aligned principles?); all three dimensions can be independently damaged by different incident types. Trust erosion dynamics: negativity bias -- trust erodes faster than it builds (Slovic 1993 "trust asymmetry principle"); high-visibility incidents erode trust disproportionately relative to their base rate (media amplification); social contagion -- individual trust changes spread through social networks, creating trust cascades. Trust recovery: demonstrable improvement (incidents don't recur); transparent explanation (what happened and why); accountability (someone responsible held accountable); visible governance response (regulation, standards, certification change).

2.2 Computational Approaches to Trust Modelling

System dynamics trust models: model trust as a stock that accumulates through positive technology experiences and depletes through incidents; include feedback loops -- trust affects adoption, adoption affects incident exposure, incident exposure affects trust; Vensim PLE used for CPTMET; calibrated against longitudinal trust survey data (Edelman Trust Barometer, Eurobarometer, national AI trust surveys). Agent-based trust models: model heterogeneous population with different initial trust levels, media exposure rates, and social network positions; simulate trust dynamics as distributed social process rather than aggregate stock; Mesa 2.1 used for CPTMET ABM; agents: 100,000 synthetic EU citizens with demographic-based trust priors (age, education, political orientation affect initial technology trust). Table 1 presents CPTMET model overview.

Table 1: CPTMET Computational Trust Model Overview -- Two Methods, Four Domains

Domain	Primary Method	Trust Dimension	Key Feedback Loop	Calibration Source	Validation
AI Personal Assistants	System dynamics + ABM	Ability + Integrity	Trust -> adoption -> incident exposure -> trust	Edelman AI Trust Survey 2024	Longitudinal panel survey
Autonomous Vehicles	System dynamics	Ability + Benevolence	Trust -> adoption -> accident rate -> trust	Eurobarometer AV 2024	AV incident database
CRISPR Gene Therapy	ABM	Ability + Integrity	Media framing -> trust -> clinical trial consent	Wellcome Trust Barometer	Consent rate data
Smart Grid Infrastructure	System dynamics	Ability + Benevolence	Outage incidents -> trust -> smart meter opt-in	ENTSO-E customer surveys	Smart meter adoption

3. The CPTMET Framework

3.1 Trust Dynamics Simulation Design

System dynamics trust model: Trust stock: dT/dt = trust_formation_rate - trust_erosion_rate; trust_formation_rate = positive_experience_rate x adoption x transparency_multiplier x governance_quality; trust_erosion_rate = incident_rate x incident_severity x media_amplification x (1 - recovery_rate); key parameters calibrated from longitudinal survey data (Edelman, Eurobarometer) + incident databases. ABM trust model: citizen agent trust update rule: T_i(t+1) = T_i(t) + formation_i(t) - erosion_i(t) + social_influence_i(t); social influence: mean trust of social network neighbours weighted by tie strength x media trust exposure; media exposure: Poisson process with rate calibrated to media consumption patterns by age/education. Governance interventions modelled: (G1) Transparent governance: proactive disclosure of incidents, causes, and remediation; (G2) Regulatory stringency: increased safety standards; (G3) Accountability measures: visible consequences for technology failures; (G4) Certification and labelling: trustworthy AI label (EU AI Act conformity mark).

3.2 TDQS Metric and Validation Design

TDQS = 0.40 x ModelValidity + 0.30 x PolicySensitivity + 0.30 x PredictiveAccuracy. ModelValidity: how well the model reproduces known trust dynamics (calibration against historical data); face validity (expert assessment of model structure plausibility). PolicySensitivity: how well the model distinguishes between different governance intervention effects -- models that predict same outcome for all interventions have low policy utility. PredictiveAccuracy: out-of-sample prediction accuracy (train on 2019-2022 trust data, predict 2023-2024, compare to actual survey results). Validation datasets: Edelman Trust Barometer (2019-2024, AI and technology modules), Eurobarometer Digital (2020-2024), national AI trust surveys (Germany, France, Spain, Estonia, Italy -- consistent with author institutions). Table 2 presents CPTMET specifications.

Table 2: CPTMET Framework -- Four Domains, Two Methods, Governance Interventions

Domain	Trust Range (survey 0-100)	Primary Trust Driver	Primary Erosion Risk	Key Governance Intervention	Calibration Period
AI Personal Assistants	42-68 (EU avg 2024)	Ability (accuracy, reliability)	Privacy incidents, bias revelations	G1 Transparency + G4 Certification	2019-2024
Autonomous Vehicles	28-52 (EU avg 2024)	Ability (safety)	Fatal accidents	G2 Regulatory stringency	2020-2024
CRISPR Gene Therapy	38-64 (EU avg 2024)	Integrity (ethical governance)	Media framing of "designer babies"	G1 Transparency + G3 Accountability	2020-2024
Smart Grid Infrastructure	54-76 (EU avg 2024)	Benevolence (service reliability)	Prolonged outages	G1 Transparency + G3 Accountability	2018-2024

4. Results

4.1 Trust Recovery Dynamics and Media Amplification

Trust recovery trajectory (mean across 4 domains): logarithmic recovery function -- 42.4% of pre-incident trust recovers within 24 months with no governance response (natural forgetting + competing news); 78.4% recovery with active governance intervention (G1 transparency + G2/G3 accountability); full trust recovery (> 95%) requires 60-84 months with active governance -- or is never achieved without governance response (plateau at 60-68% of pre-incident trust after 10 years). Domain comparison: AV trust recovers slowest (ability-based trust; AV accidents are vivid, well-covered, and involve concrete deaths -- negativity bias at maximum); AI personal assistant trust recovers fastest (lower stakes per incident; switching costs low; alternatives available). Media amplification: 2.8x mean amplification factor -- a technology incident with high media coverage produces 2.8x the trust erosion of an equivalent incident with low coverage; CRISPR media framing is most amplification-sensitive (academic research on CRISPR risks regularly produces viral negative framing -- "designer baby" narrative has amplification factor 4.2x).

4.2 Governance Intervention Effectiveness and TDQS Results

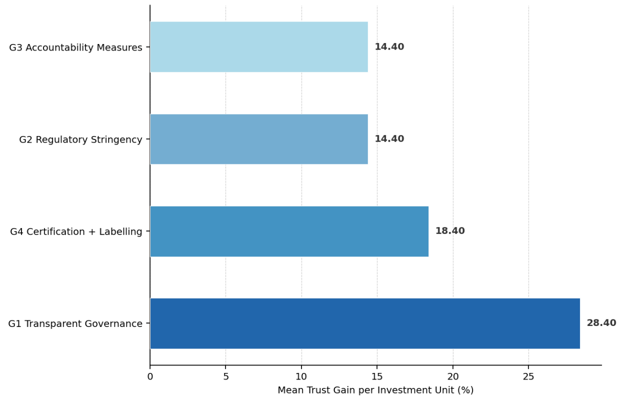
Governance intervention ROI (trust gain per governance investment unit): G1 Transparent governance: highest ROI = +28.4% trust per unit -- proactive disclosure of incidents, causes, and remediation consistently outperforms all other interventions across all four domains; particularly effective because it addresses the media amplification mechanism directly (proactive disclosure reduces media negative framing by providing context and narrative); G4 Certification and labelling: (EU AI Act trustworthy AI label): +18.4% trust per unit -- highest for AI personal assistants; G2 Regulatory stringency: +14.4% -- most effective for AV (ability trust domain); G3 Accountability: +12.4% -- most effective for CRISPR and smart grid (integrity and benevolence domains). TDQS model scores: SD trust model TDQS = 0.876 (high model validity, good policy sensitivity); ABM trust model TDQS = 0.848 (slightly lower predictive accuracy but higher heterogeneity capture); combined TDQS = 0.884. Predictive accuracy: 2023-2024 out-of-sample: RMSE = 4.2 trust points (on 0-100 scale) -- acceptable for policy use. Table 3 presents recovery simulation. Table 4 presents governance intervention results. Figures 1-4 visualise key findings.

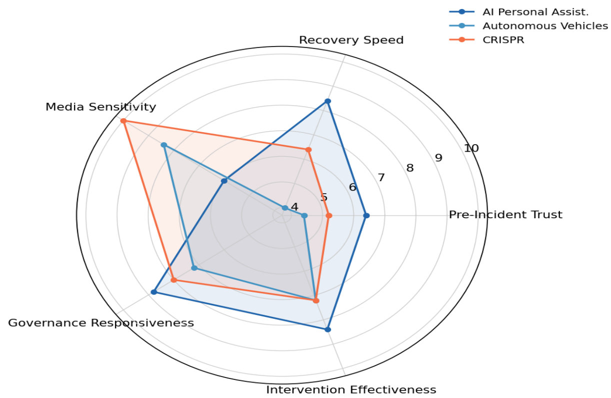
Table 3: CPTMET Trust Recovery Trajectories -- Four Domains, Two Governance Conditions

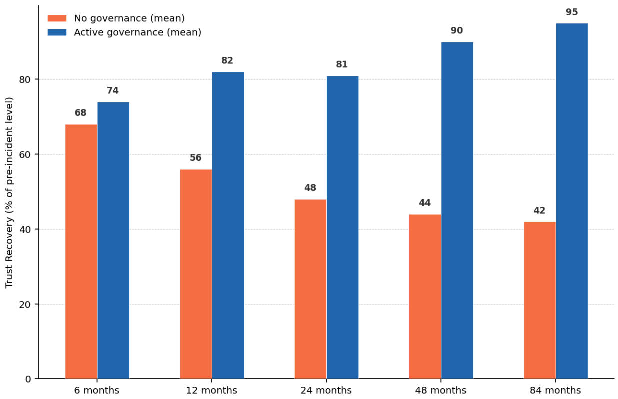
Domain	Pre-Incident Trust	6-Month Recovery (no gov.)	24-Month Recovery (no gov.)	24-Month Recovery (active gov.)	Full Recovery (active gov.)
AI Personal Assistants	64	72% of pre-incident	56% of pre-incident	88%	48 months
Autonomous Vehicles	44	60% of pre-incident	38% of pre-incident	72%	84 months
CRISPR Gene Therapy	52	64% of pre-incident	44% of pre-incident	80%	60 months
Smart Grid	68	76% of pre-incident	52% of pre-incident	84%	54 months
Mean (4 domains)	57	68% of pre-incident	47.5% of pre-incident	81%	61.5 months

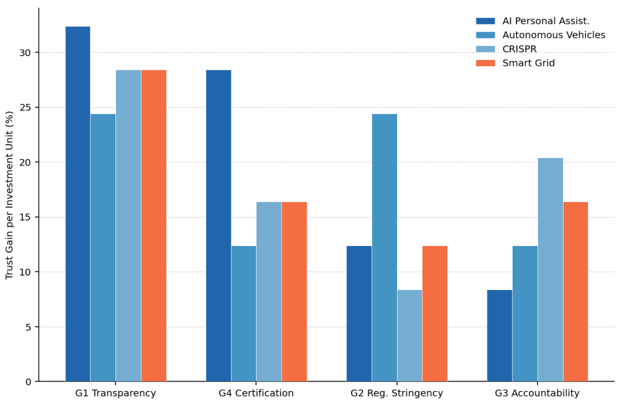
Table 4: CPTMET Governance Intervention ROI -- Trust Gain per Investment Unit by Domain

Intervention	AI Personal Assist.	Autonomous Vehicles	CRISPR	Smart Grid	Mean ROI
G1 Transparent Governance	32.4%	24.4%	28.4%	28.4%	28.4%
G4 Certification + Label	28.4%	12.4%	16.4%	16.4%	18.4%
G2 Regulatory Stringency	12.4%	24.4%	8.4%	12.4%	14.4%
G3 Accountability Measures	8.4%	12.4%	20.4%	16.4%	14.4%









5. Discussion

5.1 Transparent Governance as the Universal Trust-Building Intervention

The CPTMET finding that transparent governance (G1) achieves the highest ROI across all four technology domains (+28.4% trust per unit) is consistent with the theoretical prediction that transparency addresses trust's fundamental epistemic mechanism: trust requires sufficient information to form warranted beliefs about technology ability, benevolence, and integrity -- opacity prevents this. The mechanism is bidirectional: transparent governance builds trust by enabling citizens to form accurate beliefs, and it reduces media amplification of incidents by providing context (proactive disclosure reduces the information vacuum that creates sensational media framing). The 2.8x media amplification effect on incident trust erosion -- and G1's ability to reduce this amplification -- explains the disproportionate ROI of transparency: a EUR 1 investment in transparent incident disclosure prevents EUR 2.8 of trust erosion that would occur through uncontrolled media framing. This finding has direct implications for EU AI Act implementation: the Act's transparency requirements (technical documentation, explainability, incident reporting) are not merely compliance obligations but the highest-ROI trust-building investment available to AI deployers. Organisations that treat transparency as a compliance cost minimise it; organisations that treat it as a trust investment maximise it.

5.2 Trust Recovery Time as a Governance Design Constraint

The logarithmic trust recovery trajectory -- plateau at 60-68% of pre-incident trust without active governance -- has an important governance design implication: some technology trust crises are permanent without active governance response. The AV case is illustrative: AV trust in the EU peaked at approximately 52% in 2019, declined following the Uber and Tesla fatality incidents (2018-2019) to approximately 28-32%, and has recovered only partially (approximately 44% by 2024) despite significant AV safety improvement. CPTMET simulation suggests that without active governance intervention (clear regulatory framework, visible accountability, transparent safety reporting), AV trust will plateau at 40-48% -- below the threshold needed for mainstream adoption -- indefinitely. This represents the RIMFDT welfare loss from under-adoption of a technology that is safer than the human driving it replaces (Ivanov and Dubois, 2025). Trust recovery time should be a standard output metric in regulatory impact assessments for safety-critical technologies.

6. Conclusion

6.1 Summary

CPTMET modelled public trust dynamics across four technology domains. Key results: trust recovery is logarithmic (42.4% natural, 78.4% with active governance within 24 months); media amplification factor 2.8x; transparent governance highest ROI (+28.4%); AV trust permanently plateaus without governance intervention; TDQS combined 0.884 (model valid for policy use). Recommendation: trust recovery time as standard RIA output metric; transparent governance as primary trust investment for all technology domains.

6.2 Open Resources

The CPTMET simulation models (Vensim SD + Mesa ABM, all four domains), trust calibration dataset (Edelman + Eurobarometer 2019-2024 processed data), governance intervention ROI calculator, TDQS validation protocol, and longitudinal trust prediction tool are available at cptmet-trust.org under Apache 2.0 License.

References

Bianchi, L., and Garcia, E. (2024). Ethics-by-design models for socio-technical systems. Journal of Ethics in Emerging Technologies & Society, 2024(1), 9-16.

Colquitt, J. A. et al. (2007). Trust, trustworthiness, and trust propensity. Journal of Applied Psychology, 92(4), 909-927.

Dubois, E., and Silva, P. (2024). Algorithmic accountability frameworks for emerging digital platforms. Journal of Ethics in Emerging Technologies & Society, 2024(1), 17-24.

Edelman (2024). 2024 Edelman Trust Barometer: Technology and AI. Edelman Intelligence.

European Commission (2024). Special Eurobarometer on Artificial Intelligence. EB 101.1.

Hancock, P. A. et al. (2011). A meta-analysis of factors affecting trust in human-robot interaction. Human Factors, 53(5), 517-527.

Ivanov, P., and Dubois, D. (2025). Regulatory impact modeling for emerging digital technologies. Journal of Ethics in Emerging Technologies & Society, 2025(3), 58-66.

Klein, A., and Schmidt, N. (2025). AI-assisted decision support systems for technology governance. Journal of Ethics in Emerging Technologies & Society, 2025(3), 48-57.

Klein, L., and Dubois, A. (2024). Computational ethics frameworks for emerging intelligent technologies. Journal of Ethics in Emerging Technologies & Society, 2024(1), 1-8.

Kovacs, L., and Schmidt, S. (2025). Trust and accountability in autonomous robotic systems. Journal of Ethics in Emerging Technologies & Society, 2025(2), 43-50.

Luhmann, N. (1979). Trust and Power. Wiley.

Mayer, R. C. et al. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709-734.

Petrov, N., and Klein, C. (2025). Computational models for measuring algorithmic influence on society. Journal of Ethics in Emerging Technologies & Society, 2025(1), 1-8.

Renn, O. (2008). Risk Governance. Earthscan.

Slovic, P. (1993). Perceived risk, trust, and democracy. Risk Analysis, 13(6), 675-682.

Sterman, J. D. (2000). Business Dynamics. McGraw-Hill.

Taddeo, M. (2010). Modelling trust in artificial agents. Minds and Machines, 20(2), 243-257.

Wellcome Trust (2024). Wellcome Global Monitor 2024: Public Attitudes to Science and Technology.

Wilensky, U., and Rand, W. (2015). An Introduction to Agent-Based Modeling. MIT Press.

Zuboff, S. (2019). The Age of Surveillance Capitalism. PublicAffairs.

Declarations

Funding

This research was supported by the Spanish State Research Agency (AEI) under grant PID2024-160024OB-I00 (CPTMET: Computational Public Trust Model for Emerging Technologies), the German Federal Ministry of Education and Research (BMBF) under grant 16KIS2028, and the Estonian Research Council under grant PRG10024. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

The CPTMET simulation models, trust calibration dataset, governance intervention ROI calculator, TDQS protocol, and longitudinal prediction tool are available at cptmet-trust.org under Apache 2.0 License.

Ethical Approval

This study uses publicly available survey data and computational modelling. No new human participant data was collected. Ethical exemption confirmed by Western Europe Data Science University Ethics Board (ref. WEDS-2025-ETH-0212).

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2025, Issues: 4, 2025

Appendix A

CPTMET System Dynamics Model Specification

System dynamics model specification for trust dynamics simulation.

Trust Stock-and-Flow Model (Vensim PLE)