



**Funded by the
European Union**

EuropeanCity2	
Project number:	101178170
Project name:	European City Squared: Computational Social Science Simulation for Democracy
Topic:	HORIZON-CL2-2024-DEMOCRACY-01-06
Type of action:	HORIZON-RIA
Starting date of action:	01/01/2025
Project duration:	36 months
Project end date:	31/12/2027
Deliverable number:	D1.3
Deliverable title:	Integrated Risk, Quality, and Privacy Frameworks
EC document version:	Ver1
WP number:	WP1
Lead beneficiary:	1-AU
Main author(s):	Kristoffer Nielbo (AU)
Internal reviewers:	Luis Razo (EISM)
Nature of deliverable:	R
Dissemination level:	PU
Delivery date from Annex 1:	30 Jun 2025 (M6)
Actual delivery date:	3 Jul 2025

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor European Research Executive Agency (REA) can be held responsible for them.

Contents

1	Executive Summary.....	3
2	Introduction and Scope.....	3
3	Risk Management Framework.....	4
4	Risk Register and Updated Mitigation Strategy	6
5	Quality Assurance Process	9
6	Data Governance and Strategy	11
7	Ethical Considerations and Safeguards	13
8	Roles, Responsibilities, and Oversight.....	16
9	Continuous Monitoring and Evolution.....	17

1 Executive Summary

This document presents the Integrated Risk, Quality, and Privacy Framework for the European City Squared (EC2) project. It outlines the strategies, responsibilities, and safeguards that ensure the project meets its scientific objectives while adhering to the highest standards of risk management, research quality, and data protection.

The EC2 project addresses the urgent need for democratic innovation through computational social science, agent-based simulation, and quantum-inspired decision models. In doing so, it operates within a complex technical, ethical, and regulatory environment. This framework ensures that risks are proactively identified, managed, and mitigated; that project outputs maintain scientific rigor and integrity; and that all data processes comply with GDPR, FAIR principles, and ethical research standards.

Key components of the framework include:

1. A structured risk management process covering technical, operational, reputational, and policy-related risks, with clearly defined mitigation measures and responsibilities.
2. Quality assurance protocols for simulation design, data management, and deliverable production, ensuring outputs meet academic, technical, and regulatory expectations.
3. A comprehensive privacy and data governance strategy, with safeguards for personal data, the use of synthetic populations, and secure information handling aligned with EU legislation.
4. Ethical oversight mechanisms addressing algorithmic fairness, responsible use of AI and language models, and protection of participants in experimental settings.

This document builds upon the commitments outlined in the EC2 Grant Agreement, including Deliverable D1.2 (Data Management Plan), and reflects the consortium's commitment to transparency, accountability, and responsible research. It serves as a living framework, subject to continuous monitoring and refinement throughout the project lifecycle.

2 Introduction and Scope

This document constitutes Deliverable D1.3 for the European City Squared (EC2) project: the Integrated Risk, Quality, and Privacy Framework. It provides a detailed operational guide for identifying, managing, and mitigating project risks; ensuring scientific and technical quality; and safeguarding privacy and ethical standards in data handling.

The scope of this framework includes:

- Risk management structures covering technical, operational, reputational, and policy-related risks.
- Quality assurance processes for simulation development, data management, and deliverables.
- Data governance policies ensuring compliance with GDPR, FAIR principles, and ethical research standards.

This deliverable fulfills the requirements set out in the EC2 Grant Agreement (Grant No. 101178170) and aligns with the Horizon Europe expectations for integrated governance in research and innovation actions. It supports the consortium's obligations for project risk management, quality control, and GDPR-compliant data handling during the 36-month implementation period (2025–2027).

The framework complements Deliverable D1.2 (Data Management Plan) and other project governance documents. It applies to all project partners and work packages, ensuring that risks are systematically monitored, quality is maintained, and data handling meets EU regulatory requirements.

This document will be reviewed and updated as needed throughout the project lifecycle, with significant revisions incorporated into future management deliverables or consortium decisions.

3 Risk Management Framework

The EC2 project operates at the intersection of computational social science, quantum social choice theory, agent-based simulation, and real-world democratic experimentation. This interdisciplinary scope introduces significant technical, operational, and political uncertainties, which require proactive and project-specific risk management.

This framework defines how EC2 identifies, monitors, and mitigates risks, emphasizing technical realism, stakeholder alignment, and scientific integrity. It complements the project’s Data Management Plan (D1.2) and integrates lessons from the first six months of implementation.

3.1 Risk Identification Process

Risks are identified through:

- Iterative technical evaluations of simulation development, including:
 - Quantum-inspired algorithm development and export to classical ABM.
 - Gravitas metric calibration.
 - Active Inference and cognitive agent behavior modeling.
- Feedback from Aarhus and QuantumBasel partners regarding real-world constraints on pilot implementation.
- Continuous assessment of external factors, including public interest, political developments, public trust dynamics, and regulatory changes.

EC2 categorizes risks into:

- *Technical Risks*: Algorithm integration failures, quantum model misalignment, LLM-induced bias in synthetic populations.
- *Operational Risks*: Delays in deliverable production, partner coordination challenges, resource constraints.
- *Sociopolitical Risks*: Public misunderstanding of democratic mechanisms, low pilot participation, ideological resistance.
- *Data and Ethics Risks*: GDPR non-compliance, privacy breaches, misuse of synthetic data, algorithmic unfairness.
- *Scientific Impact Risks*: Inability to validate simulations, unrealistic outputs undermining policy recommendations.

3.2 Risk Assessment

Risks are scored using a project-specific matrix:

Score	Likelihood Definition	Impact Definition
1	Rare or theoretical	Negligible project disruption
2	Unlikely but plausible	Minor delays or technical rework
3	Possible under current conditions	Moderate disruption or reputation impact
4	Likely without intervention	Major technical failure or stakeholder loss
5	Almost certain	Severe project-wide consequences

Risk Level is calculated as Likelihood x Impact. Risks scoring ≥ 12 are flagged as high priority.

3.3 Mitigation Strategy

EC2 mitigation approaches are tailored to:

- Reduce technical complexity through modular design (quantum layer decoupled from runtime ABM).
- Control LLM influence via human-in-the-loop validation, prompt constraint engineering, and scenario testing.
- De-risk pilots through early citizen engagement and leveraging Danish concepts of “samfundssind” to build trust.
- Align with GDPR by prioritizing synthetic populations and minimizing real data use in simulations.
- Protect scientific credibility through transparent communication of model limitations and simulation assumptions.
- Apply a structured AI risk management approach across all work packages utilizing LLMs or AI-supported models, ensuring documentation, risk assessment, and human oversight consistent with EC2’s AI governance and ethical commitments.

3.4 Roles and Responsibilities

Risk management is distributed but coordinated:

- Project Coordinator (AU) + Scientific Coordinator (EISM): Central oversight, escalation authority.
 - Bi-weekly review of risk status, mitigation enforcement.
- Operational Board: Meets every 6 weeks to review risk status, mitigation enforcement.
- WP Leads: Technical and operational risk ownership within work packages.
- Ethics Lead (EISM) + Ethics Oversight Board: Oversight of data governance, privacy, and fairness risks.
- Pilot Site Partners (Aarhus, QuantumBasel): Real-world risk monitoring, stakeholder management.

Emerging high-priority risks are escalated to the Operational Board and, if unresolved, to the European Commission.

3.5 Monitoring and Adaptation

Risk management is embedded in:

- Bi-weekly coordination meetings.
- Technical work package milestone reviews.
- Pilot preparation sessions and in-person meetings with Aarhus and Basel stakeholders.
- Deliverable quality reviews.

The Risk Register (Section 4) is a living document, updated as:

- Technical realities evolve.
- Societal conditions change.
- New ethical concerns emerge, particularly regarding AI and quantum-informed algorithms.

This pragmatic, adaptive approach ensures that EC2 remains scientifically rigorous, politically viable, and compliant with European standards for research integrity and democratic innovation.

4 Risk Register and Updated Mitigation Strategy

Risk Description	Likelihood (1–5)	Impact (1–5)	Risk Level (L x I)	Work Packages	Proposed Mitigation
Limitations of quantum computing	2 (Low)	4 (High)	8 (Med)	WP3	Rely primarily on current, accessible technology (e.g., IBM Q). In addition, we have a confirmed quantum strategy that aligns with current quantum tech.
Integration challenges of LLMs	4 (High)	4 (High)	16 (High)	WP4–6	Flexible frameworks, pilot testing. We have limited the scope of implementation and identified the LLM models and approach.
Data privacy / GDPR compliance	2 (Low)	4 (High)	8 (Med)	WP4–8	Data management plan, legal audits. The DMP plan is now more developed and more mature than at the proposal stage.
Low stakeholder engagement	3 (Med)	3 (Med)	9 (Med)	All WPs	Dedicated management team, structured engagement strategy. We have implemented an Operational Management Board + Bi-weekly meetings between Coordinator and Scientific Coordinator.
Dependency on external data	3 (Med)	4 (Med)	12 (Med)	All WPs	Synthetic data creation technology has improved since our grant proposal, and we have a clearer understanding of data needs from Aarhus and QuantumBasel. This remains a high risk, but we have downgraded it based on the above factors.
No/Low adoption of policy outputs	3 (Med)	4 (High)	12 (High)	WP9	Early engagement with policymakers, actionable recommendations. This remains a significant risk, but we have significant commitment and interest from the Project Officer and EU officials in Brussels who have expressed their commitment.
Failure to meet project milestones	3 (Med)	3 (Med)	9 (Med)	All WPs	Active project monitoring, adaptive management, contingency plans. At M6, everything appears to be running smoothly and on-schedule.

Risk Description	Likelihood (1–5)	Impact (1–5)	Risk Level (L x I)	Work Packages	Proposed Mitigation
Intellectual property disputes	2 (Low)	4 (High)	8 (Med)	All WPs	Clear IP agreements, early identification of rights. No significant changes here.
Funding or resource constraints	2 (Low)	3 (Med)	6 (Low)	All WPs	Budget management, coordinator-level resource buffer. No significant changes here.
Technological obsolescence risk	2 (Low)	3 (Med)	6 (Low)	WP6–8	Market scanning, R&D investment, agile development cycles. No significant changes here, perhaps even more risk now, given the rapid pace of technological development.
Ideological extremism influencing perception	2 (Low)	4 (High)	8 (Med)	WP10	Strong anti-ideological stance in all communications. No significant changes here, although the apparent threat of authoritarianism seems higher.
Complexity of scientific ideas reduces accessibility	3 (Med)	4 (High)	12 (High)	All WPs	"Keep it simple" messaging, focus on influential communicators. We now have more clarity about the project details and key concepts, such as the Danish " <i>samfundssind</i> ," which should make communication easier.
Low public understanding of project concepts	3 (Med)	3 (Med)	9 (Med)	WP8–10	Public awareness campaigns, emphasis on trusted communicators. See previous risk and mitigation strategy.
Regulatory challenges in real-world application	3 (Med)	4 (High)	12 (High)	WP9	Early engagement with regulators and ethics boards. No significant changes from the proposal stage here. Active monitoring necessary.
Simulation realism or technical model limitations	4 (High)	4 (High)	16 (High)	WP4–6	Strong validation against real-world data, iterative refinement. Given the additional conceptual clarity we have, this threat seems slightly less likely.
Low sustainability or limited project impact post-grant	3 (Med)	3 (Med)	9 (Med)	WP10	Long-term sustainability planning, commercialization strategies. Given our tapping into the Danish concept of " <i>samfundssind</i> " and the marketability of Denmark as "the happiest country in the world," we expect stakeholders to fully capitalize on project outputs.

Additional Risks Identified in the EC2 Proposal and updated mitigation approach:

1. Potential for integration of quantum computing to amplify existing biases or introduce errors in simulation outcomes.
 - Mitigation: To minimize this risk, EC2 has deliberately chosen to decouple quantum computing from the live operation of the classical agent-based simulation. Instead of maintaining a real-time interface between quantum systems and the simulation engine, quantum models will be executed separately, with their outputs exported as static mathematical algorithms or precomputed influence weights. These outputs are then imported into the classical simulation as fixed parameters.
 - This approach contains potential risks by ensuring that any computational anomalies or biases introduced by quantum processes are confined to the preprocessing phase. If issues arise, they will be detectable and correctable prior to integration, with no risk of cascading errors during active simulation runs. This design choice significantly reduces the likelihood of quantum-derived errors distorting simulation outcomes.
2. Complexity of combining altruism and self-interest within agent-based models may lead to skewed outputs or unintended social dynamics.
 - Mitigation: Given our identification of “samfundssind” (a Danish concept referring to an individual’s conception of the common good) as the governing principle of our simulation, and our consortium partner’s access to real data from the city of Aarhus, we feel less concerned about this risk. If Danish culture has been able to strike this balance in such a remarkable way, there is no reason to believe we cannot at least equal mimic it. Nevertheless, the risk still exists and we are aware of it.
3. Resistance to adoption of unfamiliar democratic mechanisms due to lack of public trust or misunderstanding.
 - Mitigation: we believe the project’s adoption of the well-known Danish concept of “samfundssind” will be helpful in this regard. Our goal of identifying the voting rule that leads to this “emergent altruism” in Denmark and communicating it to the people of Denmark has a significant head start because of “samfundssind”. Similarly, with Denmark as a kind of social, cultural, and political benchmark of success, we believe we will reduce the risk of resistance and increase the probability of receptivity in other cities and regions.
4. Misinterpretation of quantum-enhanced simulation outcomes leading to flawed policy recommendations.
 - Mitigation: This remains a significant risk, somewhat mitigated by the previous points above, and by the fact that we are developing a rigorous “simulation” buffer between theory and real-world practice. In addition, the participation of Aarhus, with its population size, is an ideal location to pilot and test the algorithms that emerge from the quantum and classical simulations.
5. Potential unintended consequences described metaphorically as “scientific disasters” like communism.
 - Mitigation: no significant changes here. This risk remains a serious concern. We will do everything we can to communicate in a measured and conservative way, transparent about the risks that come with social and political innovation.
6. Pilot study results may be influenced by local factors, limiting generalizability.
 - Mitigation: The project design intentionally addresses this risk by focusing on algorithmic abstraction from the outset, particularly at the quantum and mathematical modeling level. Rather than tailoring voting mechanisms or decision-making models solely to local conditions, EC2 aims to extract generalized, system-level principles—what we refer to as the “algorithm of social and cultural success”—through these pilots.

- This involves validating our models using context-specific data from Danish and Swiss locality, while explicitly ensuring that the underlying voting rules or decision algorithms are formulated in a transferable, scalable manner. By grounding the approach in formal, system-agnostic logic (such as Quadratic Voting and quantum-informed fairness mechanisms), the project reduces dependence on local idiosyncrasies and enhances the potential for broader application across diverse democratic settings.
7. Stakeholder cooperation could be hindered by bureaucratic or institutional resistance.
- Mitigation – This risk remains relevant but is substantially reduced by the strategic composition of the consortium and the nature of existing partnerships. In Aarhus, the Department of Democratic Innovation—responsible for advancing governance—is a formal project partner. Their explicit mandate to explore democratic experimentation, combined with a demonstrated institutional commitment to innovation, significantly lowers the likelihood of internal resistance.
 - Similarly, the collaboration with QuantumBasel involves direct engagement with organizational leadership and decision-makers. This top-level alignment ensures that project objectives are integrated into institutional priorities from the outset, reducing bureaucratic barriers and facilitating cooperation. While external factors can never be fully controlled, the project is well-positioned to mitigate institutional resistance through these established, high-level partnerships.

5 Quality Assurance Process

Ensuring the scientific, technical, and operational quality of EC2 outputs is critical to the project's success, given the complexity of integrating agent-based simulation, quantum-inspired voting mechanisms, language models, and real-world democratic pilots. The EC2 consortium applies a multi-layered quality assurance (QA) framework that reflects the interdisciplinary nature of the project while addressing the specific risks associated with its experimental design.

This section outlines the processes in place to safeguard the quality, reliability, and validity of project deliverables, simulation outputs, data handling, and pilot activities.

5.1 Simulation and Model Quality Control

EC2's core research output is a modular simulation platform that integrates:

- Classical agent-based models (ABM) with cognitive realism through Active Inference principles,
- Quantum-informed voting weights derived via pre-processing,
- Gravititas metrics for influence sensitivity,
- LLM-generated synthetic populations and preference distributions.

Given the high stakes of simulating democratic processes, the following QA mechanisms are applied:

- Technical Validation:
 - All simulation modules undergo standalone and integrated testing to ensure:
 - Correct mathematical instantiation (e.g., preference aggregation mechanisms function in a mathematically rigorous way),
 - Consistency with theoretical models (e.g., Quadratic Voting and Quantum Quadratic Voting logic is preserved post-integration),
 - No unintended interactions between modules (quantum, LLM, ABM).
- Empirical Calibration:
 - Synthetic populations are benchmarked against available real-world data from Aarhus and Basel to ensure demographic and behavioral plausibility.
- Peer Review of Algorithms:

- New voting rules, Gravitas computations, and preference collapse mechanisms are subject to internal peer review by independent technical partners within the consortium.

5.2 Data Quality and LLM Output Control

EC2 uses synthetic datasets and LLM-generated agent profiles to reduce privacy risks while ensuring simulation diversity and realism. To maintain data quality:

- **Controlled Prompt Engineering:**
 - LLM prompts for agent initialization are standardized to minimize bias, ensure reproducibility, and capture the necessary socio-demographic diversity.
- **Human-in-the-Loop Oversight:**
 - LLM outputs for sensitive aspects (e.g., preference framing, belief distributions) are reviewed by the EC2 social science team to prevent distortions.
- **Bias Testing:**
 - Simulation scenarios include tests for algorithmic fairness, ensuring no unintended systemic advantages emerge for specific agent sub-groups.
 - Adversarial testing methodologies are applied in simulation environments to identify vulnerabilities, edge cases, or unexpected AI behaviors before LLM-generated outputs influence agent interactions or public-facing components.

5.3 Deliverable Review and Scientific Integrity

EC2 applies structured, multi-stage quality controls for all project deliverables to ensure scientific accuracy, compliance with project objectives, and alignment with Horizon Europe standards.

Internal Review Process

The deliverable review process follows a formal, consortium-wide protocol:

- **Initial Planning (Minimum Six Weeks Prior to EC Submission Deadline):**
 - The Lead Beneficiary responsible for the deliverable identifies and secures at least two internal reviewers from different beneficiary organizations.
 - The Lead Beneficiary informs the relevant Work Package Lead (WPL) of the selected reviewers, copying the internal reviewers.
- **First Internal Review (Latest Six Weeks Prior to Submission Deadline):**
 - The Lead Beneficiary circulates a first draft of the deliverable to the internal reviewers, copying the WPL and Project Coordinator.
 - Internal reviewers provide structured feedback, returning the draft with comments to the Lead Beneficiary, copied to the WPL and Coordinator.
- **Second Internal Review:**
 - The Lead Beneficiary incorporates feedback and circulates an updated draft to the internal reviewers, WPL, and Coordinator.
 - Internal reviewers conduct a second round of review, returning further comments.
- **WPL Review and Approval:**
 - The Lead Beneficiary integrates final reviewer feedback and sends the updated draft to the WPL for formal approval, copying the internal reviewers and Coordinator.
 - The WPL provides any final comments to the Lead Beneficiary, with all parties copied.
- **Coordinator Final Review (Minimum One Week Prior to Submission Deadline):**
 - The Lead Beneficiary submits the final draft to the Project Coordinator for final review, copying the WPL and internal reviewers.
 - The Coordinator provides any final comments to the Lead Beneficiary.
- **Submission to EC (Latest Three Working Days Prior to Submission Deadline):**
 - The Lead Beneficiary sends the final, formatted version of the deliverable (Word file) to Aarhus University RSO (Bojana), copying the WPL, internal reviewers, and Coordinator.

- The RSO performs a final formal check (e.g., cover page, formatting) and submits the final PDF version via the EC SyGMa portal.
- Post-Submission Documentation:
 - After submission, the RSO uploads the deliverable to the EC2 private workspace (SharePoint).
 - Public deliverables may also be made available via the EC2 project website.

Note: The Lead Beneficiary is responsible for maintaining an updated Table of Contents and for submitting the deliverable email to the RSO with high priority, using the subject line:

“Project acronym (EC2): DX.X – ready for submission to EC”

Version Control and Traceability

- A central, secure repository (SharePoint) tracks all document versions, review feedback, and approvals.
- All versions, comments, and approval stages are archived to ensure transparency, accountability, and traceability.

This comprehensive, standardized process guarantees that EC2 deliverables meet high scientific, technical, and formal quality standards before submission to the European Commission.

5.4 Pilot Activity Quality Controls

Real-world pilots in Aarhus and Basel present unique risks to project credibility and output quality:

- Stakeholder Co-Design:
 - Pilot interventions are co-designed with local stakeholders to ensure relevance, feasibility, and social acceptability.
- Iterative Prototyping:
 - Voting mechanisms (e.g., Quadratic Voting and QQV) are introduced in stages, with controlled trials before broader application.
- Empirical Feedback Loops:
 - Pilot outcomes are analyzed to calibrate simulation models, ensuring, if possible, a virtuous cycle of theoretical and empirical refinement.

5.5 Continuous Improvement and Adaptive QA

Given the project’s complexity and evolving technological landscape, EC2’s QA approach is adaptive:

- Every 6 weeks there is an Operational Board meeting to review technical and scientific progress, flagging emerging quality concerns.
- Simulation results are compared against evolving theoretical benchmarks and pilot data to maintain validity.
- Lessons learned are incorporated into subsequent simulation iterations, deliverable production, and stakeholder communications.

Through this integrated, project-specific QA framework, EC2 ensures that its technical innovations, data processes, and public-facing outputs meet the highest standards of scientific, ethical, and operational quality.

6 Data Governance and Strategy

EC2 applies a rigorous, multi-layered data governance framework that prioritizes legal compliance, scientific integrity, and public trust. The project strictly adheres to the European Union’s General Data Protection Regulation (GDPR) and operationalizes ethical data stewardship through the implementation of FAIR principles (Findable, Accessible, Interoperable, Reusable).

Data protection oversight is led by Aarhus University, acting as the designated Data Manager for the consortium, as formalized in the Data Management Plan (D1.2) and Data Specification. Data storage, processing, and access control are handled through the AU-ERDA, AU-SIF and UCloud infrastructure, which is where necessary is ISO/IEC 27001-certified, ensuring alignment with European and national research integrity standards, see Data Specification.

6.1 GDPR Compliance and Personal Data Minimization

EC2 prioritizes personal data minimization across all project activities:

- The majority of simulation and experimental outputs rely on synthetic populations and artificial datasets, reducing the use of real-world personal data.
- Where personal data is required (e.g., pilot evaluations, surveys, stakeholder interviews):
 - Informed consent procedures are implemented, supported by Participant Information Sheets and transparent privacy notices.
 - Pseudonymization or anonymization is performed prior to any data analysis or storage.
 - All personal data is stored exclusively within ISO/IEC 27001-certified Storage and Virtual Research Environments.

These measures ensure full compliance with GDPR and Danish national data protection requirements.

6.2 Synthetic Data Generation and Use

EC2 applies advanced synthetic data methodologies to enhance realism while safeguarding privacy:

- LLM-driven synthetic data generation produces realistic agent profiles, belief systems, and preference distributions. Where EC2 integrates externally developed Large Language Models qualifying as General Purpose AI (GPAI) under the EU AI Act, the consortium verifies provider compliance with transparency, technical documentation, and risk mitigation obligations.
- Synthetic datasets are calibrated to reflect demographic, economic, and political characteristics relevant to Aarhus, Basel, and broader European contexts.
- No direct reproduction of identifiable individuals or sensitive real-world information is permitted.
- Data generation processes are:
 - Transparent and documented.
 - Subject to validation against empirical reference data where available.
 - Audited to ensure alignment with the consortium's Data Management Plan (D1.2) and GDPR requirements.

Synthetic data reduces dependence on external, real-world data sources, while supporting experimental realism and inclusivity.

6.3 Data Stewardship and FAIR Principles

EC2 ensures responsible data management through full implementation of FAIR principles:

- Findable: DOIs are assigned via Zenodo and Metadata records are linked from project outputs to public repositories.
- Accessible: Data are stored in AU-SIF and AU-ERDA with metadata consistent with the DCAT v3 schema. Open data sets will be deposited at Zenodo and sensitive data will be accessible upon request.
- Interoperable: Use of standard data formats (CSV, JSON) and Metadata vocabularies consistent with RDF standards (DCAT).
- Reuseable: Open access datasets licensed under CC-BY 4.0. Data and metadata will be maintained for at least five years after project conclusion. Documentation of processing and provenance is provided per dataset (see Data Specification, e.g., LLM_Synthetic_Preferences_provenance.md).

6.4 Data Lifecycle Management

Data governance applies across the full data lifecycle:

- **Collection:** Only data necessary for scientific or pilot objectives is collected, with a preference for synthetic or anonymized datasets.
- **Processing:** Secure handling procedures are followed, with clear processing records maintained in accordance with GDPR.
- **Storage:** Data is stored in AU-ERDA, AU-SIF UCloud Virtual Research Environment, ensuring physical and digital security.
- **Sharing:** Controlled, legally compliant data sharing is permitted within the consortium and, where appropriate, through public repositories for synthetic or anonymized datasets.
- **Archiving/Deletion:** Data is securely archived or deleted post-project in line with contractual obligations, data protection laws, and the Data Management Plan and Data Specification.

6.5 Ethical Oversight of Data Use

- The consortium's Ethics Advisory Board (EAB) provides independent oversight on data governance and privacy matters.
- The Ethics Lead (EISM) monitors compliance with ethical standards, including:
 - GDPR adherence.
 - Responsible AI and LLM deployment.
 - Avoidance of algorithmic bias or misuse of synthetic data.
- Data governance processes are reviewed periodically and updated based on project evolution, stakeholder feedback, and regulatory changes.

Through this integrated data governance framework, EC2 ensures that its technical innovations, simulation outputs, and real-world interventions are conducted responsibly, ethically, and in full compliance with European legal and research standards.

7 Ethical Considerations and Safeguards

The EC2 project recognizes that its work on democratic experimentation, advanced simulation, and AI integration raises significant ethical considerations. These include safeguarding fundamental rights, ensuring algorithmic fairness, protecting public trust, and complying with legal and regulatory frameworks, particularly the GDPR and emerging EU AI governance standards.

This section reflects the ethical safeguards implemented across EC2, incorporating the requirements detailed in Deliverable D11.1 (POPD-H Requirements), D11.2 (Data Protection Measures), and the formal governance role of the Ethics Advisory Board (EAB) as outlined in the EAB Mandate and Memorandum of Understanding.

7.1 Responsible AI and Language Model Integration

EC2 integrates Large Language Models (LLMs) into its simulation environment for agent profile generation, preference modeling, and simulated discourse.

- LLM use is governed by:
 - Controlled prompt engineering, minimizing bias and ensuring contextual relevance.
 - Human-in-the-loop validation, with outputs reviewed by domain experts prior to integration.
 - Bias detection protocols, applied systematically to monitor for distortions or unfair outcomes.
- EC2 aligns with the risk-based principles of the EU AI Act, ensuring transparency, accountability, and explainability of AI components.

The EAB retains oversight over all AI deployments, including LLM outputs that influence democratic processes or simulated decision-making.

7.2 Algorithmic Fairness and Democratic Integrity

- Voting mechanisms, including Quadratic Voting and quantum-preprocessed Gravitas weights, undergo:
 - Peer-reviewed design processes.
 - Testing for bias, representational fairness, and resilience to manipulation.
- The Gravitas metric, operationalized through Fisher Information and quantum formalisms, is evaluated to prevent systemic bias or concentration of influence.
- The Ethics Advisory Board reviews all algorithms affecting agent influence, preference collapse, and vote aggregation.

7.3 Protection of Pilot Participants

- *Informed Consent*: All pilot participants are provided with clear, accessible Participant Information Sheets detailing their rights, the experimental nature of the project, and data protection measures.
- *Inclusion of Vulnerable Groups*: Vulnerable individuals may only participate incidentally and with EAB oversight to ensure additional safeguards are in place.
- *Data Protection Impact Assessment (DPIA)*: A formal DPIA is conducted prior to each pilot, ensuring that risks related to privacy, fairness, and participant welfare are identified and mitigated.
- *Anonymization*: Pilot data is pseudonymized or anonymized prior to analysis, with all personal data processed in compliance with GDPR and stored exclusively within UCloud’s certified environment.

7.4 Public Trust, Accessibility, and Democratic Accountability

- Communication strategies prioritize:
 - Simplified, accessible messaging explaining complex concepts like Gravitas, quantum social choice, and simulation-based decision models.
 - Engagement with trusted local communicators, leveraging concepts such as “samfundssind” to foster community understanding and trust.
- The consortium maintains strict neutrality, avoiding political or ideological alignment in both scientific outputs and public communications.

7.5 Ethics Oversight Structure

- The Ethics Advisory Board (EAB):
 - Operates independently, providing external oversight on all ethics, privacy, AI, and pilot-related issues.
 - Meets at least twice per year, issuing formal ethics reports to the Project Coordinator and Steering Committee.
 - Holds authority to recommend changes to project processes or pause activities that present unresolved ethical risks.
- The Ethics Lead (EISM) ensures:
 - Continuous monitoring of project activities for ethical compliance.
 - Coordination of internal reviews, DPIAs, and EAB consultations.
 - Integration of updated ethical standards into project operations as EU AI and data governance frameworks evolve.

7.6 Compliance with the EU AI Act

The EC2 project proactively aligns its use of Artificial Intelligence (AI) with the risk-based framework established by the European Union’s Artificial Intelligence Act (AI Act), formally adopted in 2024. While EC2’s core research involves agent-based simulation and democratic experimentation, several AI components integrated into the project qualify as high-risk applications under the AI Act. These include:

- LLM-driven synthetic population generation, used to create realistic but non-identifiable agent belief systems, preference distributions, and demographic characteristics.
- LLM-assisted proposal framing and scenario generation, used to model narrative influence and test information exposure effects within simulations.

- Preference evolution and communication modeling, supported by LLM outputs in controlled simulation environments, with human oversight.
- Integration of quantum-preprocessed influence weights and Gravitas metrics, where AI techniques assist in validating preference aggregation processes and fairness metrics.

While most AI activity occurs within the controlled boundaries of EC2's synthetic simulations, certain project components intersect with AI Act obligations, particularly:

- The generation of synthetic populations and behavioral profiles, which, while anonymized, model real-world demographic characteristics.
- Simulation outputs with potential downstream influence on public decision-making processes or policy recommendations.
- Human-facing components in project pilots (e.g., citizen assemblies, participatory processes) where AI-influenced scenarios or preference models are presented for deliberation or feedback.

In response to these identified risks and regulatory expectations, EC2 implements the following AI Act compliance measures:

Risk Management and AI Documentation (see previous sections)

- A structured AI risk management system is embedded across all work packages utilizing LLMs or AI-supported models.
- All AI development, including LLM output generation, preference modeling, and synthetic population construction, is documented, with detailed records maintained for:
 - Training datasets.
 - Prompt engineering methodologies.
 - Validation protocols and bias assessments.

Human Oversight and Explainability

- Human-in-the-loop oversight applies to all AI-generated content influencing agent behavior, preference distributions, or narrative framing within the simulations.
- If and when LLM outputs are used for public-facing activities (e.g., proposal framing in pilots), explainability requirements are applied to ensure participants understand AI-generated content.
- Bias detection and fairness validation are conducted before AI outputs are integrated into simulations or presented in public-facing project activities.

General Purpose AI (GPAI) Considerations

- Where EC2 integrates externally developed or sourced LLMs classified as General Purpose AI (GPAI) under the AI Act:
 - The consortium verifies that GPAI providers comply with transparency, technical documentation, and systemic risk mitigation obligations.
 - Internal risk assessments, adversarial testing, and documentation of downstream uses are maintained, in alignment with GPAI requirements.

Ethics Oversight and Continuous Monitoring

- The Ethics Advisory Board (EAB) holds authority to review all AI use cases within EC2, with specific attention to:
 - High-risk AI deployments.
 - Fairness and bias mitigation.
 - Alignment with GDPR and fundamental rights protections.
- AI-related risks and compliance measures are reviewed bi-annually, with updates provided in annual Ethics Reports submitted to the Operational Board.

Through these integrated safeguards, EC2 ensures that its AI components, including LLMs and AI-supported democratic simulations, are designed and deployed responsibly, ethically, and in full alignment with European legal and regulatory standards, including the AI Act.

8 Roles, Responsibilities, and Oversight

The EC2 Operational Board understands that effective risk management, quality assurance, and data governance in EC2 depend on clearly defined roles, responsibilities, and oversight structures. Given the project's complexity—combining computational social science, AI integration, quantum-preprocessed algorithms, and real-world democratic pilots—the consortium has developed the a governance model.

This section outlines the specific responsibilities of key partners and structures that ensure risks are identified, mitigated, and continuously monitored; quality standards are upheld; and privacy, ethical, and legal obligations are met across all project activities.

8.1 Project Coordination and Risk Oversight

- Project Coordinator (Aarhus University) with the help of Scientific Coordinator (EISM):
 - Holds overall responsibility for project-wide risk management.
 - Monitors deliverable production, compliance, and milestone achievement.
 - Ensures escalation of critical risks to the Operational Board or, where necessary, to the European Commission.
 - Coordinates project-level responses to emerging scientific, technical, or operational risks.
 - Conducts bi-weekly reviews of project status, technical progress, and emerging risks.

8.2 Work Package Leadership

Each Work Package (WP) Lead is accountable for:

- Day-to-day risk identification and mitigation within their respective WP.
- Integration of ethical, privacy, and data governance measures into all activities.
- Reporting risks, quality concerns, or ethical issues to the Operational Board and Ethics Lead as they arise.

8.3 Operational Board and Governance

The EC2 Operational Board:

- Meets every 6 weeks to review project status, technical progress, and emerging risks.
- Coordinates cross-WP mitigation strategies and quality control interventions.
- Oversees the internal deliverable review process, ensuring all outputs meet scientific, technical, and regulatory standards before submission.

8.4 Ethics Advisory Board (EAB)

- The EAB operates as an independent, external oversight body, mandated to:
 - Review AI deployments, quantum-informed algorithms, and simulation outputs for fairness, transparency, and compliance with the EU AI Act and GDPR.
 - Monitor pilot activities for adherence to ethical standards, protection of participants, and responsible democratic experimentation.
 - Provide bi-annual ethics reviews and formal annual reports.
- The Ethics Lead (EISM):
 - Coordinates internal ethics risk assessments.
 - Oversees data protection compliance and implementation of the Data Management Plan (D1.2).
 - Facilitates integration of EAB recommendations into project operations.

8.5 Data Governance and Privacy

- The Data Manager (Aarhus University):
 - Ensures GDPR compliance across all data processing activities.
 - Provides guidance on lawful data handling, data minimization, and synthetic data use.

- Coordinates DPIAs prior to pilot activities involving personal data.
- All consortium partners:
 - Adhere to data governance protocols as outlined in D1.2.
 - Participate in ongoing data quality, privacy, and ethics training.

8.6 AI Oversight

- Work Packages utilizing AI, including LLMs and synthetic population generation (WP4–6):
 - Implement AI risk management processes and documentation requirements.
 - Ensure human evaluation of AI outputs.
 - Verify compliance of external AI providers with AI Act obligations.
- The EAB retains authority to audit AI use, assess systemic risks, and recommend interventions if high-risk AI applications present unresolved ethical or legal concerns.

8.7 Escalation Procedures

If risks, quality issues, or ethical concerns cannot be resolved within a Work Package or by the Operational Board:

- The issue is escalated to the Project Coordinator for immediate intervention.
- For unresolved or project-wide concerns, the matter is brought to the Project Officer.
- In cases involving legal, regulatory, or ethical non-compliance, the European Commission is informed in accordance with contractual obligations.

9 Continuous Monitoring and Evolution

The EC2 project recognizes that risk management, quality assurance, and data governance are not static processes but require continuous adaptation as technical, ethical, and regulatory landscapes evolve. The integration of quantum-informed algorithms, AI-supported models, and real-world democratic experimentation introduces inherent uncertainties that necessitate iterative monitoring and responsive governance.

This section outlines how the consortium ensures that risk mitigation strategies, quality standards, privacy protections, and ethical safeguards remain effective and are adapted in response to emerging challenges or changes in project scope.

9.1 Ongoing Risk Monitoring

- Risk identification and assessment are embedded in regular project operations, including:
 - Bi-weekly Operational Board meetings, where WP leads and the Project Coordinator review technical progress, project milestones, and emerging risks.
 - Deliverable production workflows, which include structured internal review processes that assess not only quality but risk exposure associated with project outputs.
 - Simulation development cycles, where technical and scientific risks (e.g., algorithmic performance, AI bias, quantum-classical integration) are evaluated iteratively.
- The Risk Register (Section 4) is maintained as a living document, updated quarterly, with additional revisions triggered by:
 - Identification of new risks.
 - Escalation of risk severity or likelihood.
 - Changes to project design or external factors (e.g., regulatory updates, political developments affecting pilots).

9.2 Adaptive Quality Assurance

- Quality control mechanisms are reviewed and refined based on:
 - Simulation testing outcomes.
 - Synthetic data validation results.

- Feedback from pilot stakeholders in Aarhus, Basel, and other engagement activities.
- The Operational Board coordinates quality assurance improvements across technical work packages, ensuring that lessons learned inform future deliverable production and simulation development.

9.3 Data Governance Evolution

- The Data Management Plan (D1.2) is updated as required to reflect:
 - Changes in data processing activities.
 - New AI model integrations or external General Purpose AI (GPAI) obligations under the EU AI Act.
 - Evolving privacy risks or legal requirements identified by the Data Protection Officer (AU) and Ethics Lead (EISM).
- The consortium tracks emerging data protection standards, AI governance developments, and European Commission guidelines to ensure continued compliance.

9.4 Ethics Oversight and AI Risk Updates

- The Ethics Advisory Board (EAB) conducts bi-annual reviews of:
 - AI system deployments.
 - Algorithmic fairness.
 - Pilot participant protections.
 - Broader research ethics across the project.
- Ethics risk assessments are adapted based on:
 - New technical deployments (e.g., LLM refinements, quantum algorithm updates).
 - Changes in project activities, including pilot interventions.
 - Regulatory developments, particularly regarding AI, data protection, and democratic governance.
- Annual Ethics Reports, submitted to the Steering Committee, include updated risk profiles, recommended safeguards, and compliance updates.

9.5 Stakeholder and Pilot Feedback Integration

- Real-world feedback from pilot activities is integrated into:
 - Simulation refinement.
 - Risk mitigation strategies.
 - Communication practices designed to enhance public trust and understanding.
- The project applies an iterative co-design approach with pilot partners to ensure interventions remain socially acceptable, technically sound, and ethically robust.

9.6 Living Framework Commitment

This deliverable (D1.3) is treated as a living document and will be updated as needed throughout the project lifecycle. Substantial revisions will be documented and shared across the consortium, with critical updates communicated to the European Commission in accordance with reporting obligations.