

Identifying AI Risks for Non-Human Life in Urban Spaces

MUSTAFA BURAK ERKOÇAK, Politecnico di Torino, Italy
KAAN RIZA AKARÇAY, Politecnico di Torino, Italy
BARIŞ TAN ÜNAL, Politecnico di Torino, Italy

Urban artificial intelligence (AI) systems are mostly engineered for human needs, leaving their underlying algorithms blind to the complex urban ecosystems and non-human life that coexist in our shared environment. While existing research remains divided between traditional AI safety and theoretical ecological frameworks, the field lacks a clear, systematic methodology to map AI-related harms across the full range of non-human life in urban environments. To address this gap, we present an end-to-end framework for evaluating AI impacts on urban spaces. We first compile a structured dataset of 257 urban AI systems. Building upon these, we compile a list of 262 non-human urban organisms and categorize them. We then design an LLM-driven pipeline to determine the specific risks each organism could suffer when exposed to each AI system. Finally, we analyze the different risks emerging across organism categories and release our analytical pipeline and dataset as an open-source contribution.

Additional Key Words and Phrases: Urban AI, Ecosystems, Non-Human Life, AI Safety, Large Language Models, Risk Assessment

1 Introduction

Urban artificial intelligence (AI) systems are rapidly expanding, yet they are mostly engineered to optimize human efficiency and convenience [11]. However, cities are not simply human habitats; they are shared environments where multiple organisms live. Non-human animals—ranging from companion animals and street dogs to urban wildlife—live directly within and are affected by these technological ecosystems [5]. Furthermore, animals are not merely "natural capital"; they are active co-creators of urban value. Urban pollinators, scavengers, plants, etc. perform critical functions that support biodiversity and ecosystem functioning. Ultimately, the lives of these organisms possess value that transcends their economic utility, making the detection of AI-induced harm against them crucial [5].

Despite their value and ecological importance, non-humans are systematically marginalized in technology design. Standard frameworks suffer from "speciesism," a paradigm that ignores the needs and risks of some non-humans [5, 14]. Contemporary AI risk assessment frameworks share this problem. While recent work attempts to transition toward the idea of multi-species cities, the practical field of AI safety remains mostly anthropocentric (centered around human values). In tackling the impact of AI on urban environments, related work has failed to provide a concrete methodology to map AI capabilities against non-human organisms.

To address this gap, we present an end-to-end framework for evaluating AI impacts on urban spaces. Specifically, we contribute the following:

- (1) We compile a structured list of AI systems in cities.
- (2) We compile a list of non-human urban organisms in cities, and categorize them.

- (3) We design a pipeline that determines the risks that each organism could experience from each AI system.

- (4) We analyze the different risks emerging across non-human organism categories.

2 Related Work

Next, we briefly review prior work regarding the theoretical AI risk assessment frameworks, empirical measurement of AI harms, and structured risk elicitation.

AI Ethics to Non-Human Life Historically, AI risk assessment and ethical frameworks have been mostly anthropocentric, focusing on human impacts such as algorithmic bias, privacy, and socioeconomic inequality [16]. However, a body of literature has begun to theoretically map how AI systems impact non-human life. Coghlan and Parker established a framework that produced five risk categories for AI's impact on animals: **1) Intentional: socially condemned/illegal**, **2) Intentional: socially accepted/legal**, **3) Unintentional: direct**, **4) Unintentional: indirect**, and **5) Foregone benefits** [6]. To derive these categories, the authors adapted animal welfare scientist David Fraser's classification of anthropogenic threats to animals [9]. By applying Fraser's logic to emerging technologies, they produced these five risk categories as their study's output to map how AI inherits human-caused risk. While this framework provides a perspective for recognizing the vulnerabilities of animals to AI technologies, it remains largely theoretical and lacks the definition of a mechanism for real-world risk elicitation.

Empirical Measurement of Ecological AI Harms Beyond theory, recent research has started to empirically measure AI's impact on animals, though this work has remained constrained to digital and informational domains. For instance, Guerrero-Casado et al. highlighted how AI-generated images and videos distort public perceptions of wildlife, misdirecting conservation funding due to biases toward charismatic mammals [12]. Similarly, the introduction of ANIMALHARBENCH by Kanepajs et al. takes a step towards quantifying animal harm risks within Large Language Model (LLM) text generation [15]. By using an LLM-as-a-judge framework, they demonstrated that LLMs exhibit distinct biases and varying harm risks depending on the organisms and context. While these works successfully quantify digital threats to animals, they remain detached from the physical realities of AI systems in urban/shared environments.

Structured Risk Elicitation To effectively scale risk assessment across complex populations, researchers increasingly rely on LLM-driven evaluations. For example, Bogucka et al. utilized a state-of-the-art LLM and a structured rubric to analyze how AI harms are amplified when human identities intersect [4]. We draw direct methodological inspiration from this approach. However, their

Authors' Contact Information: Mustafa Burak Erkoçak, s338631@studenti.polito.it, Politecnico di Torino, Turin, Italy; Kaan Rıza Akarçay, s337492@studenti.polito.it, Politecnico di Torino, Turin, Italy; Barış Tan Ünal, s338210@studenti.polito.it, Politecnico di Torino, Turin, Italy.

framework relies entirely on human sociological categories, which makes it incapable of assessing risks to non-human organisms.

To sum up, previous work has failed to provide a clear, systematic map of AI-related risks across the full-range of non-human life in urban environments.

3 Methodology

We come up with a pipeline to map the dimensions of AI risks to non-human organisms (Fig. 1). We first construct representative datasets of urban organisms and AI systems, and then execute a LLM-based counterfactual auditing pipeline.

3.1 Dataset Construction

We construct two datasets to capture the a diverse list of non-human organisms and variety of representative urban AI systems.

3.1.1 Non-Human Organisms. We collected raw organism occurrence data from the Global Biodiversity Information Facility (GBIF) across eight global reference cities: Singapore, Sydney, New York City, Berlin, Santiago, Beijing, Los Angeles, and Tokyo. These cities were specifically selected to maximize global diversity across four factors: Climate, Biodiversity Realm, Urban Typology, and Technological Advancement. The following steps were specifically followed:

- (1) **Sourcing and Log-Dampened Poisson Sampling:** We collected occurrence records using exact spatial convex hulls for each city. Specifically, the complete taxonomy was extracted to establish an initial pool of the top 1000 most frequently observed organisms per location.

Raw occurrence data inherently suffers from citizen observer bias, where highly visible species (e.g. urban birds) accumulate significantly more than micro-fauna or flora. To balance the list without applying arbitrary quotas, we employed a Log-Dampened Poisson Sampling approach: First, we applied a logarithmic dampening transformation, $\log_{10}(1 + C_i)$, to the raw occurrence counts (C_i) to compress extreme click-bias. Second, we applied a taxonomic penalty by dividing a species' inclusion probability by its local Family count, $N(\text{Family}_i)$. We defined the final inclusion probability (P_i) for each candidate species as:

$$P_i = \min\left(1.0, \frac{K \cdot P(\text{Occurrence}_i)}{N(\text{Family}_i)}\right) \quad (1)$$

where K is a custom multiplier which ensures that the sum of all inclusion probabilities is equal to a target expected value (of species, per city). Here, we set the target expected value to 30 (i.e. $E[N] = \sum P_i = 30$), and scale K accordingly to satisfy this condition. This probability-based approach helped neutralize observer bias and included underrepresented macroscopic kingdoms in our list. Executing this Poisson method yielded a total of 244 sampled species across the reference cities: Santiago (39), Singapore (36), Los Angeles (30), Tokyo (30), Berlin (29), Beijing (29), New York City (28), and Sydney (23)

- (2) **Inclusion of Human-Proximate Organisms:** Due to the nature of GBIF Occurrence dataset, we manually included

nine human-proximate categories (eight popular pets and one houseplant category) into the matrix [19].

We directly included five of these categories (dogs, cats, gold-fish, rabbit, and mouse) with a single GBIF specie name since these pets exhibit a single scientific name. For the remaining four categories (parrots, turtles, hamsters, and houseplants), we included the most common local representative from each city's wild pool, querying GBIF for the top order/family representative. The repeated specie names were eliminated. As a result, 18 human-proximate organisms were collected. After merging both, a list consisting of 262 unique non-human organisms were obtained.

- (3) **Category Labeling:** We use GPT-5-mini to assign one category for each of nine dimensions to each organism. We also evaluate a tenth dimension, Social Desirability, using Gemini Pro 3.1 guided by a specialized system prompt (provided in Appendix D.3). Since this dimension requires human-perspective, we evaluated the LLM responses via majority voting among the three authors, on 50 randomly sampled organisms. The majority votes within the authors matched for 88% of the sample set. These ten dimensions (see in detail below) are derived from four criteria, which we selected for their value in risk elicitation and analysis:

1) Perceptual Detectability: This criterion captures the physical features of the organism. It derives the dimensions Scale [18] (the general volume and mass of the organism), Morphological Form [21] (the anatomical structure and shape of the organism), and Kinematics [13] (the speed and pattern of the organism's movement).

2) Spatial Intersection: This criterion defines the physical environment the organism belongs to. It derives the dimension Space [10, 17], which classifies the primary physical substrate (e.g. aquatic, terrestrial, arboreal, etc.) the organism requires for its core life functions.

3) Socio-Cultural Valuation: This criterion defines the values humans have regarding the organism. It derives the dimensions Economic Effect [2] (the valuation of the organism based on its ecosystem services or property impacts), Governance Status (the management type, or conservation type applied to the organism), and Social Desirability (the social and cultural perception of the organism by humans).

4) Sensory Vulnerability: This criterion defines the organism's ecological and biological sensitivities. It derives the dimensions Sensory Vulnerability [7] (the primary sensory modality, such as acoustic or visual, the organism relies upon to perceive its environment), Rhythm [20] (the natural temporal pattern of the organism's activity across the daily cycle), and Ecological Status [3] (how the species adapts to, exploits, or avoids human-built environments).

For a complete list of the categories in to each dimension, see Table 1. Each organism was assigned only one category for each dimension.

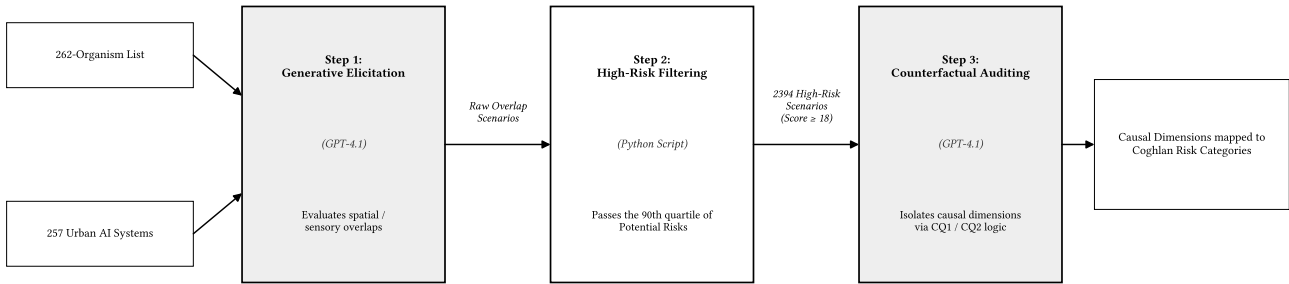


Fig. 1. **Overview of our three-step methodology for identifying the causal dimensions of AI risks to non-human organisms.** The methodology integrates two datasets—a 262-organism list and 257 urban AI use cases—into an elicitation model that produces raw spatial and sensory overlap scenarios (1). A deterministic script subsequently filters these scenarios by calculating a composite Risk Score and passing 2394 (90th percentile) high-risk instances (2) to an auditing model to find final causal dimensions via counterfactual auditing logic (3).

3.2 Pipeline Architecture

We process the list of urban organisms and the urban AI systems through a three-step pipeline to map risks according to the Coghlan risk categories.

- (1) **Generative Elicitation (GPT-4.1):** Using the system prompt given in Appendix D.1, the LLM produces the plausible risks between the AI systems and the organisms. The model drops implausible interactions (returning empty arrays) and generates risks by applying biological logic. It generates risks systematically and assigns them to one of the five Coghlan risk categories [6]: 1) *Intentional (socially condemned/illegal)*, 2) *Intentional (socially accepted/legal)*, 3) *Unintentional (direct)*, 4) *Unintentional (indirect)*, and 5) *Foregone benefits*. For each matched risk, the model outputs the necessary fields (*Risk_ID*, *AI_System_ID*, *AI_System_Name*, *Description*) and assigns a *Severity* and *Frequency* score on a Likert scale (1-7).
- (2) **High-Risk Filtering:** A deterministic Python script calculates a Risk Score for each potential risk by multiplying its *Severity* and *Frequency* [1] (yielding a maximum possible score of 49). To isolate only the most critical risks for deeper analysis, we use a 90th percentile cutoff on the distribution of Risk Scores. Put another way, we advance only the top 10% of the 23,945 total generated risks to the next step. The mathematical cutoff for this percentile emerged as a score of 18 (out of 49).
- (3) **Counterfactual Auditing (GPT-4.1 via Batch API Processing):** Using the system prompt given in Appendix D.2, the LLM performs a strict counterfactual audit on the high-risk scenarios that advanced from Step 2. It evaluates all nine biologically inherent dimensions assigned to the organism by testing them against two causal questions: whether the risk exists *because* of the specific dimension (CQ1), and whether it would still occur if the organism possessed a different value for that dimension (CQ2). For example, if an autonomous delivery rover fails to detect and collides with a domestic cat, the auditor asks: “Did this risk occur because of the cat’s small physical scale?” (CQ1 = Yes), and

“Would this collision still occur if the organism possessed a massive scale, such as a horse?” (CQ2 = No). The final output comprises of dimensions that received a “Yes” for CQ1 and a “No” for CQ2, identifying whether a specific category is the primary cause of that risk.

4 Evaluation

The goal of this evaluation was to evaluate whether providing the LLM with the organisms’ categories improves the plausibility and quality of generated AI risks compared to the outputs generated using a naive, flat risk mapping prompt. Specifically, we evaluate the LLM outputs for the Step 1 (Generative Elicitation).

4.1 Metrics

We measured performance using a Plausibility Score (1-5). Using this metric, we evaluated four distinct criteria: 1) the biological and physical possibility of the interaction/risk; 2) the relevancy of the assigned Severity and Frequency scores; 3) the accuracy of the Coghlan categorization; and 4) appropriate restraint (the model’s ability to correctly return empty arrays for impossible interactions). To quantify Inter-Rater Reliability (IRR), we calculated Exact Agreement, Adjacent Agreement (± 1 point), and Quadratic Weighted Cohen’s Kappa (κ).

4.2 Setup

We utilized blinded human evaluation. We established a test set of 30 unique AI-Organism pairs via stratified sampling: 80% (24 pairs) were positive known combinations, and 20% (6 pairs) formed a negative control group to specifically test the prompt’s resistance to hallucinating risks. For execution, we assigned the GPT-4.1 temperature strictly to 0.0 to guarantee deterministic, reproducible outputs across both the naive and structured prompts.

4.3 Execution

We executed both prompts against the 30 test pairs, extracted the text outputs, and randomly blinded them as “Method A” and “Method

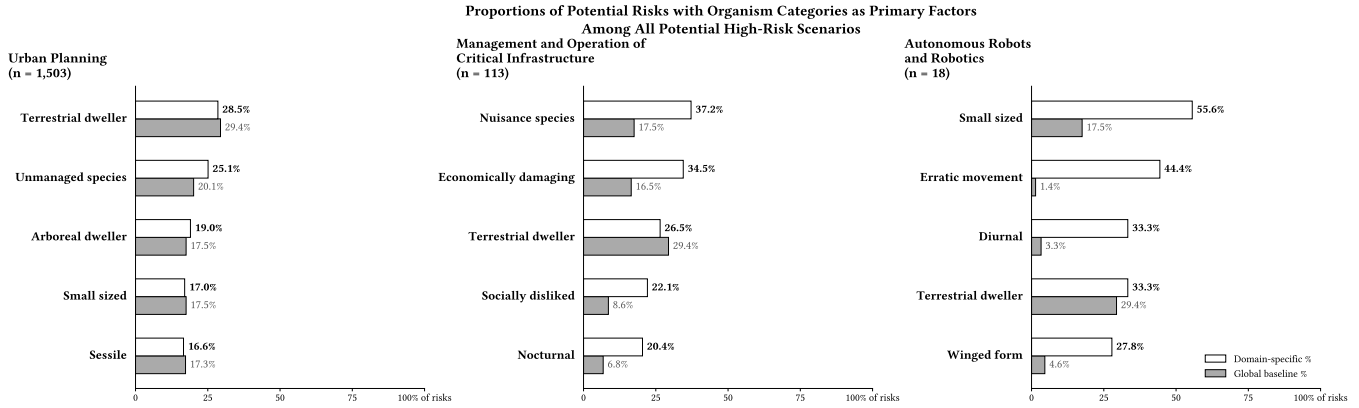


Fig. 2. **Proportional impact of organism dimensions on specific AI risk categories.** Values represent the percentage of risks within a given Coghlan risk category (e.g., Unintentional Direct) where the counterfactual algorithmic audit identified the organism’s dimension as a definitive causal factor for the harm. Because individual AI-organism interactions frequently possess multiple concurrent causal dimensions (e.g., an interaction driven simultaneously by spatial overlap and small physical scale), row and column percentages are independent and do not sum to 100%.

B” to eliminate human grading bias. Two independent authors evaluated the blinded outputs. The two raters applied the Plausibility Score to the entire output block rather than individual risk nodes.

4.4 Results

The evaluation yielded an Exact Agreement of 31.7%, an Adjacent Agreement of 73.3%, and a Quadratic Weighted Cohen’s Kappa (κ) of 0.375. The high adjacent agreement indicated that raters perceived plausibility similarly but had slightly different internal strictness on their ratings. To mathematically resolve this variance, we calculated the Mean Rater Score for each output to form an implicit consensus Gold Standard.

Upon unblinding the data, the structured pipeline significantly outperformed the naive baseline:

- **Naive Baseline Average Score:** 3.05 / 5.0
- **Structured Prompt Average Score:** 3.50 / 5.0

The structured methodology achieved a validated +14.8% **performance improvement**. The baseline score of 3.05 demonstrates that an unconstrained LLM relying purely on latent weights generates mediocre, or occasionally hallucinated risks. Conversely, constraining the model with the 10-dimension biological schema grounds to scientific plausibility.

5 Results and Discussion

The data points to a structural property of urban AI deployment: these systems are designed by humans and for human ends. Developers set the safety thresholds, define the training objectives, and choose what the model should optimize for, so the resulting algorithms reason about human priorities and stay effectively blind to the non-human organisms that share the same urban space. This anthropocentric optimization becomes measurable when our ecological risk data is cross-referenced against the risk tiers of the EU AI Act [8], an alignment we report as the Regulatory Blind Spot Index (Appendix Table 6). The pipeline assigns a higher mean ecological severity to the systems the Act labels “Limited or Low Risk” (2.82)

than to those it labels “High Risk” (2.57). The single “Prohibited” case scores higher still, but with only two instances it carries no weight. The inversion holds where the volume of systems is large, which indicates that the regulatory attention given to a system tracks its danger to people rather than its danger to ecosystems.

A second pattern reinforces this reading. Most of the audited risks fall under the two Unintentional categories of the Coghlan framework [6], with 1,758 indirect and 64 direct cases out of the 2,394 high-risk scenarios. The majority of the potential harm to urban ecosystems is therefore a consequence of omission rather than malice. Non-human life is rarely accounted for during design and development, and the resulting systems disrupt ecological baselines as a side effect of doing what they were built to do.

5.1 Domain-specific threat profiles

Because the harm originates in design omissions, developers and researchers need to treat ecological exposure as a property of the deployment domain. The impact of an AI system is not uniform across sectors, and two quantities make the differences concrete: how densely a domain produces risk, and through what physical mechanism that risk reaches an organism.

Normalizing the generated risks by the number of unique systems per domain controls for the uneven composition of the dataset (Appendix Table 4). Autonomous Robots and Robotics produces the highest density at 243.2 risks per deployed system, even though its absolute volume is small. Urban Planning generates the largest raw count (7,078 risks across 56 systems) but a moderate density of 126.4, while virtual or narrowly scoped domains such as Finance and Investment (30.2) sit at the bottom. Density and volume are different lenses, and a domain can be ecologically intensive per system without dominating the totals.

The mechanism of harm separates domains just as sharply. The Sectoral Kinetic Footprint (Appendix Table 5) reports the share of a domain’s unintentional risk that is direct, such as a physical collision, against the share that is indirect, such as habitat exclusion or

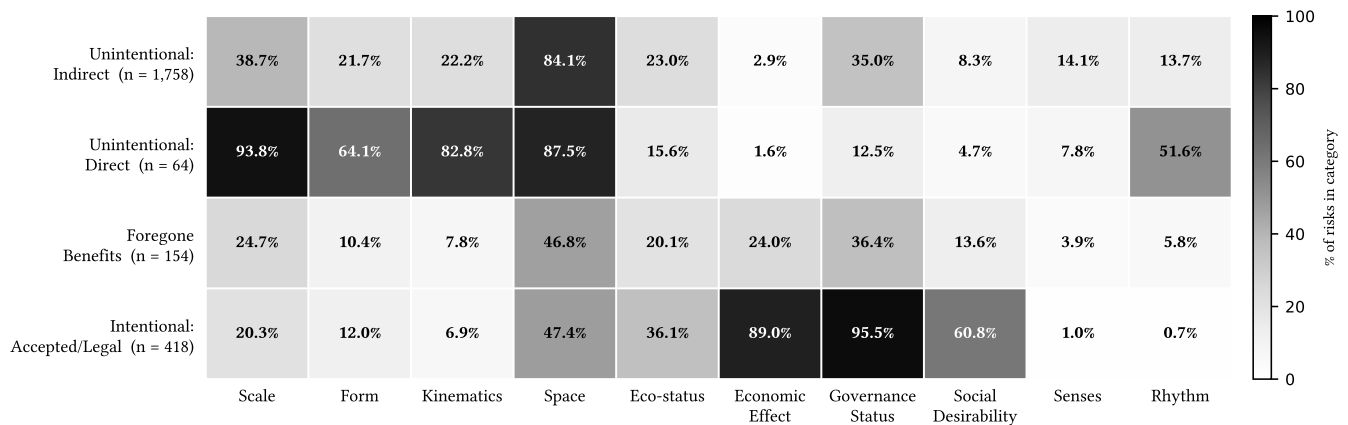
Heatmap of Causal Dimension Density Across AI Risk Categories


Fig. 3. A breakdown of the primary biological and socio-cultural traits driving AI-induced risks across three distinct deployment domains, compared against their global baseline averages. The data demonstrates how the physical nature of the AI end-system dictates ecological vulnerability. Domains reliant on physical movement (Autonomous Robotics) disproportionately endanger organisms based on physical kinematics and scale. Conversely, stationary infrastructure domains (Critical Infrastructure, Urban Planning) endanger organisms based on spatial intersection and human-assigned valuations (e.g., governance nuisance status).

acoustic disruption. Autonomous Robotics is the only domain where the two are comparable (51.64% direct against 46.30% indirect, a direct-to-indirect ratio of 1.12), because its systems operate physical vehicles moving through shared space. Domains built on virtual or highly localized software sit at the opposite end: Energy (0.02), Finance and Investment (0.04), and Smart home (0.05) generate almost no direct contact and instead harm organisms indirectly through their infrastructure and resource demands. The physical nature of the end system, not the sophistication of the model, governs the kind of ecological risk a domain produces.

5.2 Organism category-based vulnerabilities

These differences are not driven by domain alone. They are tied to the ten biological and socio-cultural dimensions of the organisms themselves. By tracing the causal source of each risk through the counterfactual (CQ) audit, we find that some categories attract disproportionate harm. Across the full audited set, terrestrial spatial intersection (`space_ter`, 29.4%) and unmanaged governance status (`gov_unmanaged`, 20.1%) are the most frequent causal traits, and at the level of whole dimensions, spatial position alone is a causal factor in 75.4% of risks, followed by governance status (45.0%) and physical scale (36.0%). A complete breakdown of trait-level and dimension-level causality appears in Appendix Tables 2 and 3.

The more useful finding is that a given dimension endangers organisms differently depending on the operational goal of the domain. Figure 3 contrasts three representative domains against the global baseline, and each one foregrounds a distinct vector of harm.

Urban Planning is dominated by a spatial-exclusion vector. Its leading causal traits are `space_ter` (28.5%), `gov_unmanaged` (25.1%), and `space_arb` (19.0%), because infrastructure expansion overlays and removes the terrestrial and arboreal habitats that sessile or ground-bound organisms depend on.

Management and operation of critical infrastructure follows a socio-cultural and economic valuation vector. Here the harm tracks human-assigned worth: `gov_nuisance` (37.2%), `econ_damaging` (34.5%), and `soc_dislike` (22.1%) lead, as systems optimize to clear or ignore organisms classed as pests. The same pattern is even stronger in Agriculture and Farming and in Environment and Sustainability, where `econ_damaging` and `gov_nuisance` together account for the majority of risk (Appendix Figure 5).

Autonomous Robots and Robotics is governed by a kinematic and scale vector, where physical parameters override governance status. Its dominant traits are `scale_small` (55.6%), `kin_erratic` (44.4%), and `rhythm_diurnal` (33.3%), because small, unpredictably moving, daytime-active animals evade collision-avoidance logic that is tuned for human-scale detection. Public and private transportation shows the same skew toward small scale and diurnal activity (Appendix Figure 9). The three domains, taken together, show that the organism categories most at risk shift predictably with what the AI system physically does.

5.3 Validation through trait-risk alignment

Mapping the Coghlan risk categories against the ten organism dimensions produces the correlation structure the framework predicts. As Figure 3 shows, organisms with particular traits are systematically affected by particular kinds of risk. Physical traits concentrate in Unintentional Direct harm, where `scale_small` (56.25%) and `space_ter` (53.12%) are the dominant causal factors, consistent with collision and crushing as the underlying mechanism. Human-valuation traits concentrate in the Intentional, socially accepted category, where `gov_nuisance` (93.78%) and `econ_damaging` (89.0%) are causal in almost every case, consistent with the deliberate removal of organisms judged to be pests. Spatial and governance traits such as `space_ter` (32.65%) and `gov_unmanaged` (25.14%) drive the

Unintentional Indirect category, consistent with habitat shading and exclusion. Because a single interaction often has several concurrent causes, the rows and columns of Figure 3 are independent and do not sum to 100%. The alignment between the theoretical risk categories and the biological taxonomy confirms the internal consistency of the two schemes and supports the pipeline's capacity to map the intersection of urban AI and ecology in a way that holds together with established risk theory.

6 Conclusion

Urban AI systems are built for humans. They optimize for our traffic, our deliveries, and our convenience. The algorithms ignore the non-human life sharing the same city blocks simply because no one told the models to look for them.

We built a pipeline to map these blind spots. We crossed 262 urban organisms with 257 AI systems. Instead of just guessing at the outcomes, we used a three-step process: a generative model drafted the raw risks, a strict mathematical filter isolated the most severe scenarios, and a counterfactual audit pinned down the exact biological or social trait that caused the danger.

The data proves that AI harm is not random. It follows strict, predictable patterns based on how a system operates. If an algorithm controls a physical vehicle or robot, it will routinely collide with small, erratic animals. The collision-avoidance sensors just do not register them. If a system manages city infrastructure or agriculture, it will systematically displace organisms that human institutions label as pests. The moving systems hit what they cannot see, and the stationary systems remove what we do not value.

AI safety has to mean more than just protecting human users. We are open-sourcing our datasets and the auditing pipeline so other researchers can use them. The immediate next step is taking these risk models out of the lab and validating them in actual city environments.

Author Positionality Statement

We are engineers and researchers working in humanitarian institutions, so our perspectives naturally reflect human priorities. We designed this framework to identify ecological risks and support non-human life in smart cities. However, we recognize that the biological and social categories we selected are still human approximations of how these organisms experience the world.

Ethical Considerations

We used Large Language Models to assign socio-cultural and ecological categories to species. Because these models train on human data, they can reproduce human biases. To reduce this risk, we designed specialized anthrozoological prompts and validated the outputs manually through majority voting. We built this dataset and framework to reduce harm by identifying the unintended impacts of algorithms on urban biodiversity. Users should not apply these tools for wildlife surveillance or ecological exploitation.

Code and Data Availability

All code and data are open source so others can reproduce our results or expand on this research. The repository includes the complete taxonomy of 262 species, the log-dampened Poisson sampling scripts, the LLM prompts, and the final labeled matrices. These materials are available at: https://github.com/kaanAkarçay/Identifying_AI_Risks_for_Non-Human_Life_in_Urban_Spaces

References

- [1] Louis Anthony (Tony)Cox Jr. 2008. What's Wrong with Risk Matrices? *Risk Analysis* 28, 2 (2008), 497–512. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1539-6924.2008.01030.x> doi:10.1111/j.1539-6924.2008.01030.x
- [2] Magda Argueta-Guzmán, Mari West, Marilia P. Gaiarsa, Christopher W. Allen, Jacob M. Cecala, Lauren Gedlinske, Quinn S. McFrederick, Amy C. Murrillo, Madison Sankovitz, and Erin E. Wilson Rankin. 2023. Words matter: how ecologists discuss managed and non-managed bees and birds. *Scientometrics* 128, 3 (Mar 2023), 1745–1764. doi:10.1007/s11192-022-04620-2
- [3] Robert B. Blair. 1996. Land Use and Avian Species Diversity Along an Urban Gradient. *Ecological Applications* 6, 2 (1996), 506–519. arXiv:<https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.2307/2269387> doi:10.2307/2269387
- [4] Edyta Bogucka, Sanja Šćepanović, and Daniele Quercia. 2026. Why AI Harms Can't Be Fixed One Identity at a Time: What 5300 Incident Reports Reveal About Intersectionality. arXiv:2604.24519 [cs.CY] doi:10.48550/arXiv.2604.24519
- [5] Daniela Borissova and Dimitar Keremedchiev. 2022. *Handbook of Management Science*. Springer Nature, Cham, Switzerland. doi:10.1007/978-3-031-17580-9
- [6] Simon Coghlan and Christine Parker. 2023. Harm to Nonhuman Animals from AI: a Systematic Account and Framework. *Philosophy & Technology* 36, 25 (2023).
- [7] EMBL-EBI. 2026. *GO:0007600 — sensory perception*. QuickGO, Gene Ontology Consortium. <https://www.ebi.ac.uk/QuickGO/term/GO:0007600> Accessed: 2026-05-03.
- [8] European Parliament and Council of the European Union. 2024. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- [9] David Fraser. 2012. A “Practical” Ethic for Animals. *Journal of Agricultural and Environmental Ethics* 25, 5 (2012), 721–746. doi:10.1007/s10806-011-9353-z
- [10] Siria Gámez and Nyeema C. Harris. 2022. Conceptualizing the 3D niche and vertical space use. *Trends in Ecology & Evolution* 37, 11 (2022), 953–962. doi:10.1016/j.tree.2022.06.012
- [11] Francesca Grasso, Stefano Locci, and Luigi Di Caro. 2025. Towards Addressing Anthropocentric Bias in Large Language Models. In *Proceedings of the 1st Workshop on Ecology, Environment, and Natural Language Processing (NLP4Ecology2025)*, Valerio Basile, Cristina Bosco, Francesca Grasso, Muhammad Okky Ibrohim, Maria Skeppstedt, and Manfred Stede (Eds.). University of Tartu Library, Tallinn, Estonia, 84–93. <https://aclanthology.org/2025.nlp4ecology-1.18/>
- [12] José Guerrero-Casado, Tamara Murillo-Jiménez, Antonio J. Carpio, Francisco S. Tortosa, and Rocío Serrano-Rodríguez. 2026. Threats to conservation from artificial-intelligence-generated wildlife images and videos. *Conservation Biology* 40, 1 (2026), e70138. arXiv:<https://conbio.onlinelibrary.wiley.com/doi/pdf/10.1111/cobi.70138> doi:10.1111/cobi.70138
- [13] Lars Haalck, Michael Mangan, Barbara Webb, and Benjamin Risse. 2020. Towards image-based animal tracking in natural environments using a freely moving camera. *Journal of Neuroscience Methods* 330 (2020), 108455. doi:10.1016/j.jneumeth.2019.108455
- [14] Thilo Hagendorff, Leonie N. Bossert, Yip Fai Tse, and Peter Singer. 2023. Speciesist bias in AI: how AI applications perpetuate discrimination and unfair outcomes against animals. *AI and Ethics* 3, 3 (Aug 2023), 717–734. doi:10.1007/s43681-022-00199-9
- [15] Arturs Kanepajs, Aditi Basu, Sankalpa Ghose, Constance Li, Akshat Mehta, Ronak Mehta, Samuel David Tucker-Davis, Eric Zhou, Bob Fischer, and Jacy Reese Anthis. 2025. What do Large Language Models Say About Animals? Investigating Risks of Animal Harm in Generated Text. arXiv:2503.04804 [cs.CY] <https://arxiv.org/abs/2503.04804>
- [16] Alexandra Sasha Luccioni and David Rolnick. 2022. Bugs in the Data: How ImageNet Misrepresents Biodiversity. arXiv:2208.11695 [cs.CV] <https://arxiv.org/abs/2208.11695>
- [17] Andrei Miljutin. 2009. Substrate utilization and feeding strategies of mammals: Description and classification. *Estonian Journal of Ecology* 58 (03 2009), 60–71. doi:10.3176/eeco.2009.1.07
- [18] Fendy Santoso, Matt Garratt, Mark Pickering, and Md Asikuzzaman. 2016. 3D-Mapping for Visualisation of Rigid Structures: A Review and Comparative Study.

IEEE Sensors Journal 16 (03 2016), 1484–1507. doi:10.1109/JSEN.2015.2498940

- [19] TGM Research. 2024. Most Common Pets Owned Worldwide 2024. TGM StatBox. <https://tgmstatbox.com/stats/most-popular-types-of-pets-owned-worldwide/> Data Source: TGM Global Pet Care Survey 2024. Accessed: 2026-06-05.
- [20] John H. Welsh. 1938. Diurnal Rhythms. *The Quarterly Review of Biology* 13, 2 (1938), 123–139. <http://www.jstor.org/stable/2808495>
- [21] Ali Zia, Renuka Sharma, Abdelwahed Khamis, Xuesong Li, Muhammad Husnain, Numan Shafi, Saeed Anwar, Sabine Schmoelzl, Eric Stone, Lars Petersson, and Vivien Rolland. 2025. A Review on Coarse to Fine-Grained Animal Action Recognition. arXiv:2506.01214 [cs.CV] <https://arxiv.org/abs/2506.01214>

Appendix A. Taxonomic Dimensions

Table 1. Taxonomic dimensions and category codes used to classify non-human organisms in the urban AI risk framework. Codes are typeset in monospace; section headers describe each dimension.

Code	Description
Physical Scale — <i>Categorized by general volumetric bounds to determine sensor threshold detectability.</i>	
scale_micro	Microscopic (< 1 mm, invisible to spatial sensors).
scale_mini	Miniature (1 mm – 10 cm; e.g., insects, moss; causes sensor noise).
scale_small	Small (10 cm – 1 m; e.g., rodents, small birds, weeds; high false-negative risk).
scale_medium	Medium (1 m – 2 m; e.g., dogs, human-sized shrubs; highly detectable).
scale_large	Large (> 2 m; e.g., mature trees, deer, livestock).
Morphological Form — <i>The geometric shape the AI's computer vision must classify.</i>	
form_bipedal	Upright / two-legged (e.g., birds on the ground, primates).
form_quadrupedal	Horizontal with limbs (e.g., cats, foxes).
form_apodal	Limbless / serpentine (e.g., snakes, worms).
form_winged	Avian / winged in flight (e.g., flying birds, bats).
form_sessile_branched	Rooted with sprawling structures (e.g., plants, trees, fungi).
form_amorphous	No distinct rigid geometry (e.g., slime molds, microbial colonies).
Kinematic Profile — <i>The predictability and speed of the organism's movement.</i>	
kin_sessile	Immobile / rooted (e.g., flora, fungi).
kin_linear	Steady, predictable trajectory (e.g., walking mammals).
kin_erratic	Unpredictable, rapid directional changes (e.g., birds, flying insects, squirrels).
Primary Spatial Niche — <i>The primary physical substrate or vertical layer the organism inhabits or requires for its core life functions. Classify based on the organism's foundational relationship to the environment, not the physical structure the organism creates.</i>	
space_aq	Aquatic: exclusively water-dwelling; requires a liquid medium for survival.
space_aq_sub	Aquatic Subterranean: occupies underwater environments beneath the surface / ground.
space_aq_ter	Aquatic Terrestrial: transitions between water bodies and land surfaces.
space_aq_sub_ter	Aquatic Subterranean Terrestrial: uses water, underground, and surface environments.
space_sub	Subterranean: exclusively occupies soil, earth, or rock below the surface level.
space_ter	Terrestrial: anchored to, rooted in, or primarily moving across the ground surface / soil.
space_sub_ter	Subterranean Terrestrial: occupies both the surface ground and the soil / earth beneath it.
space_arb	Arboreal: lives primarily on or within elevated botanical structures (e.g., branches, canopies) rather than the ground.
space_ter_arb	Terrestrial Arboreal: transitions between the ground surface and elevated botanical canopies.
space_star	Subterranean Terrestrial Arboreal: utilizes underground, surface, and elevated botanical spaces.
space_aer	Aerial: primary activities (foraging, mating) occur in the open airspace.
space_ter_aer	Terrestrial Aerial: transitions between ground-level occupation and open airspace.
Urban Ecological Status — <i>How species relate to human environments.</i>	
eco_obligate	Human-dependent.
eco_exploiter	Thrives in heavily built environments.
eco_adapter	Uses urban green spaces / edges.
eco_avoider	Cannot survive urban disruption.
Economic Effect — <i>Human valuation of ecosystem services or impacts.</i>	
econ_beneficial	Beneficial impact.
econ_neutral	Neutral impact.
econ_damaging	Damaging impact.
Economic & Governance Status — <i>Official municipal or legal standing.</i>	
gov_protected	Legally protected / conserved.
gov_cultivated	Owned / farmed / purposefully maintained.
gov_nuisance	Targeted for removal / extermination.
gov_unmanaged	Outside formal policy.
Social Desirability — <i>Public emotional categorization based on anthrozoology and conservation psychology.</i>	
soc_dislike	Actively disliked or repulsed by the general public.
soc_indifferent	Elicits little to no emotional response or awareness.
soc_save	Perceived as vulnerable and worthy of conservation efforts.
soc_love	Highly charismatic, beloved, or culturally valued.
Primary Sensory Perception — <i>Dominant mechanism used to perceive environmental stimuli. Choose the modality most continuously relied upon in ordinary sensing / foraging / orientation, not rare or specialized capabilities.</i>	
sense_magneto	Magnetoreception.

Continued on next page

Table 1 continued from previous page

Code	Description
sense_photo	Light stimulus.
sense_mechano	Mechanical stimulus.
sense_thermo	Temperature stimulus.
sense_chemo	Chemical stimulus.
sense_electro	Electrical stimulus.
sense_hygro	Humidity stimulus.
Activity Rhythm – Primary temporal pattern of organism–environment interaction across a normal 24-hour cycle. For organisms without sleep–wake behavior and with continuous physiological–environmental interaction, prefer continuous.	
rhythm_diurnal	Day-active.
rhythm_nocturnal	Night-active.
rhythm_crepuscular	Dawn / dusk-active.
rhythm_continuous	Always on / a-rhythmic.

A Appendix B: Domain-Specific Risk Distributions

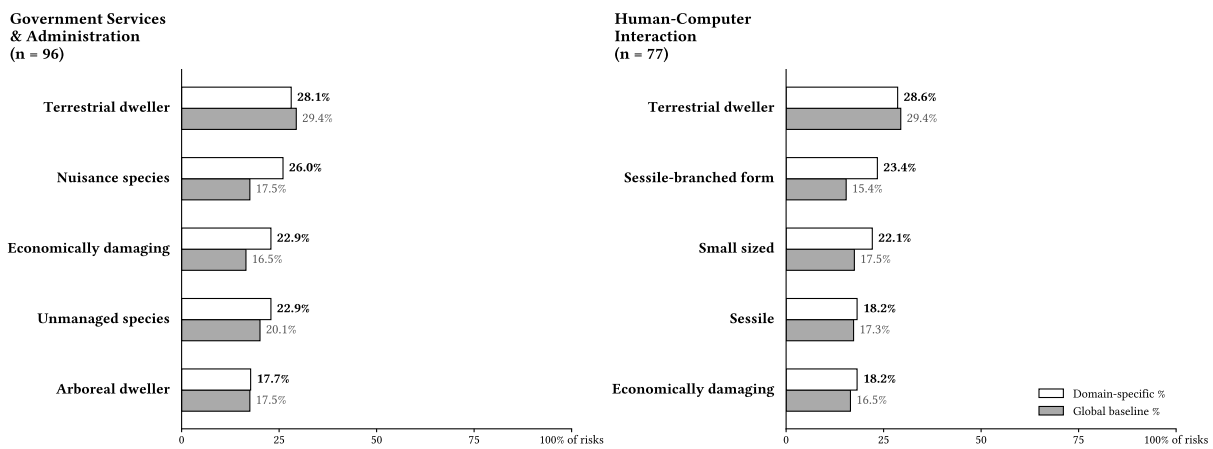


Fig. 4. Comparison of domain-specific risk percentages against the global baseline for Government Services & Administration (n = 96) and Human-Computer Interaction (n = 77)[cite: 2].

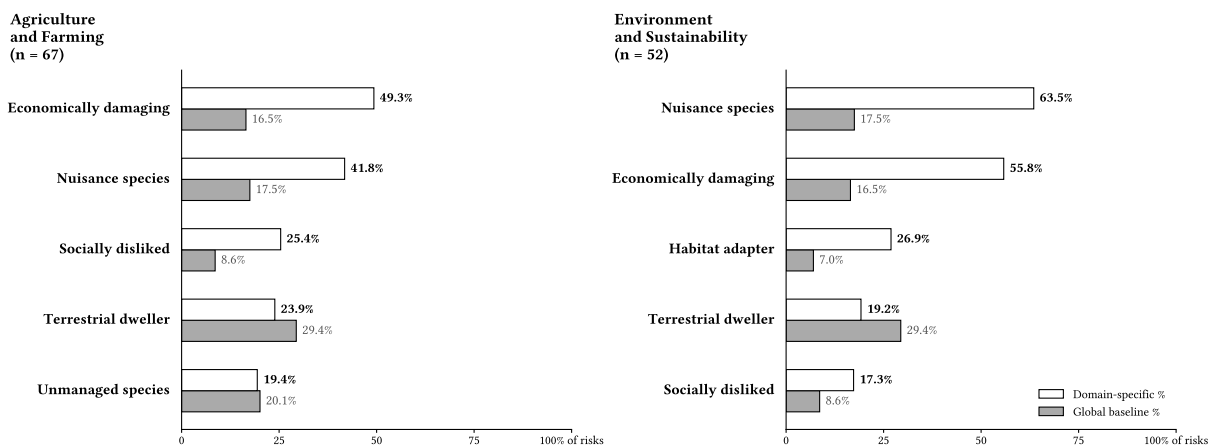


Fig. 5. Comparison of domain-specific risk percentages against the global baseline for Agriculture and Farming (n = 67) alongside Environment and Sustainability (n = 52)[cite: 3].

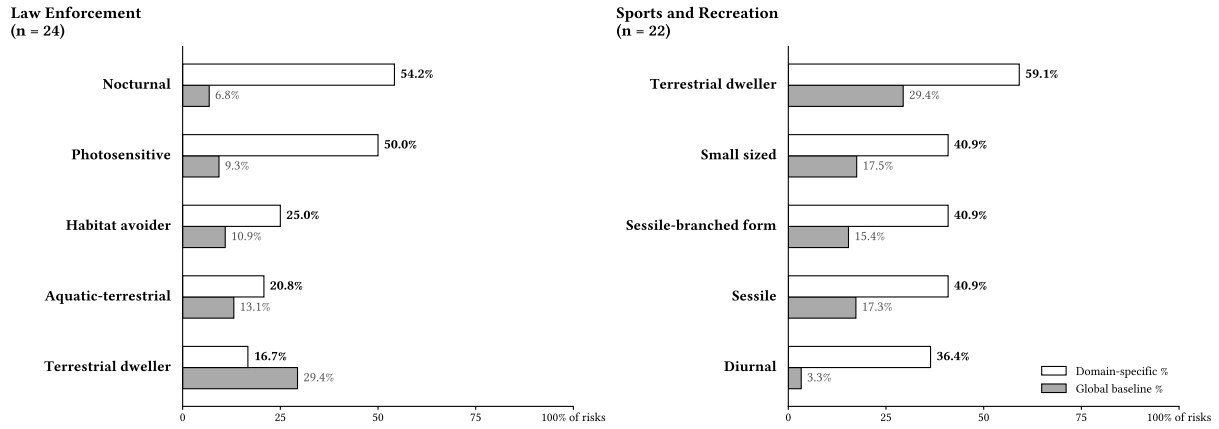


Fig. 6. Comparison of domain-specific risk percentages against the global baseline for Law Enforcement ($n = 24$) and Sports and Recreation ($n = 22$)[cite: 4].

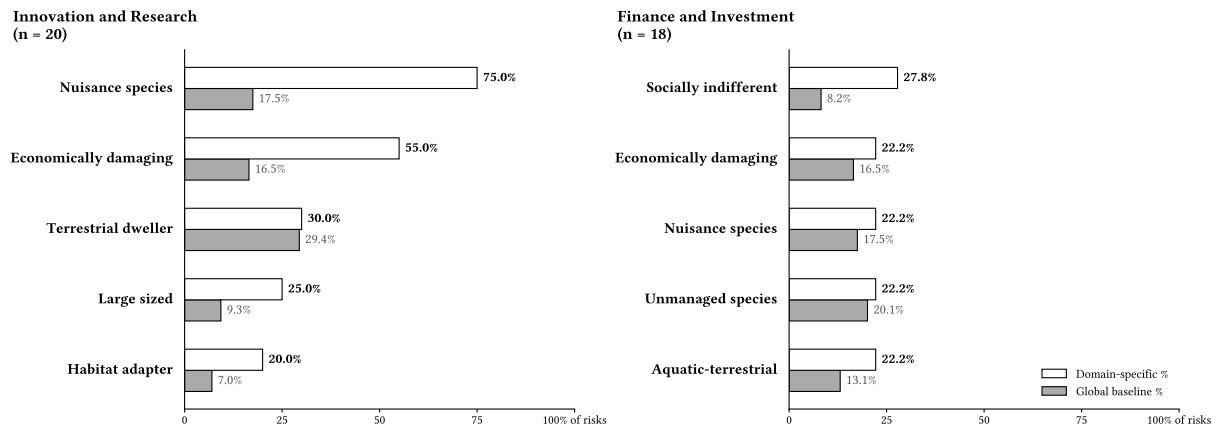


Fig. 7. Comparison of domain-specific risk percentages against the global baseline for Innovation and Research ($n = 20$) and Finance and Investment ($n = 18$)[cite: 5].

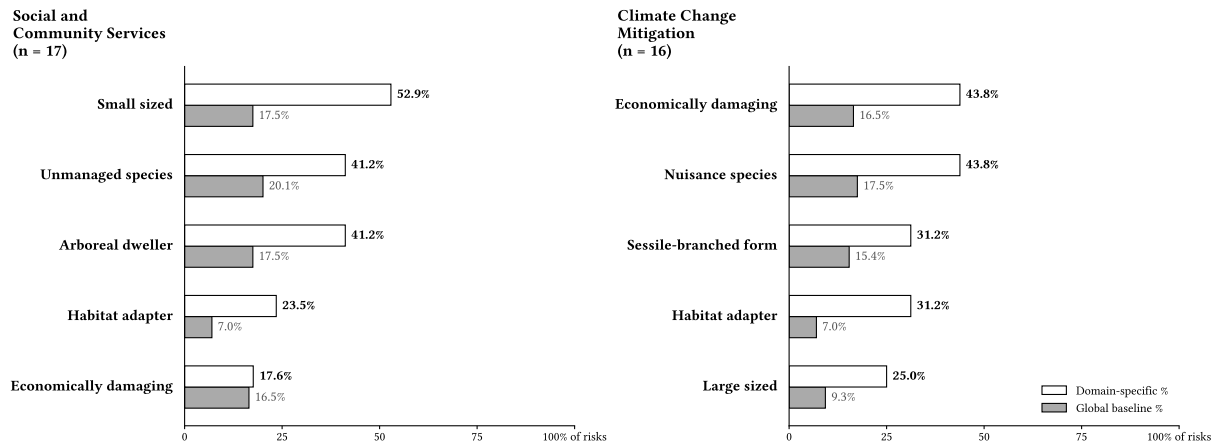


Fig. 8. Comparison of domain-specific risk percentages against the global baseline for Social and Community Services ($n = 17$) and Climate Change Mitigation ($n = 16$)[cite: 6].

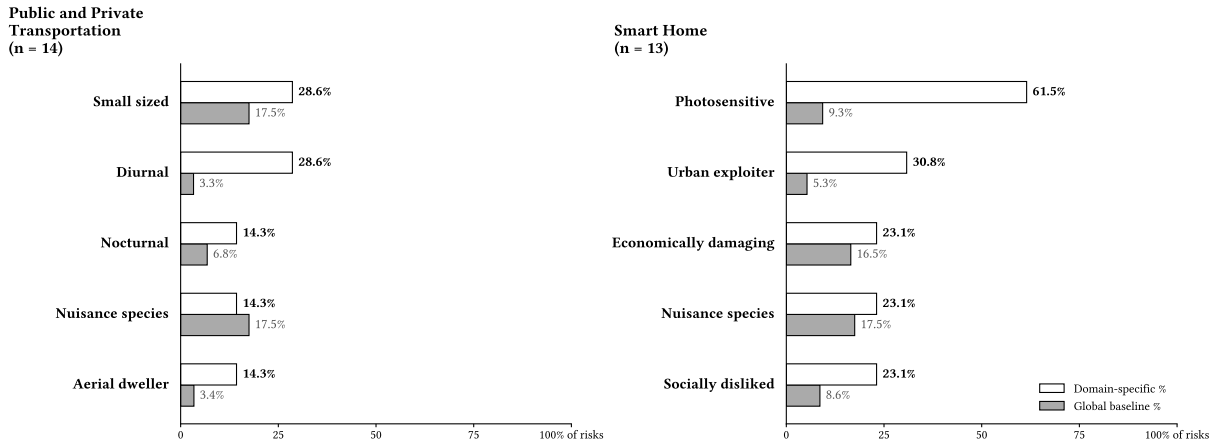


Fig. 9. Comparison of domain-specific risk percentages against the global baseline for Public and Private Transportation ($n = 14$) and Smart Home applications ($n = 13$)[cite: 7].

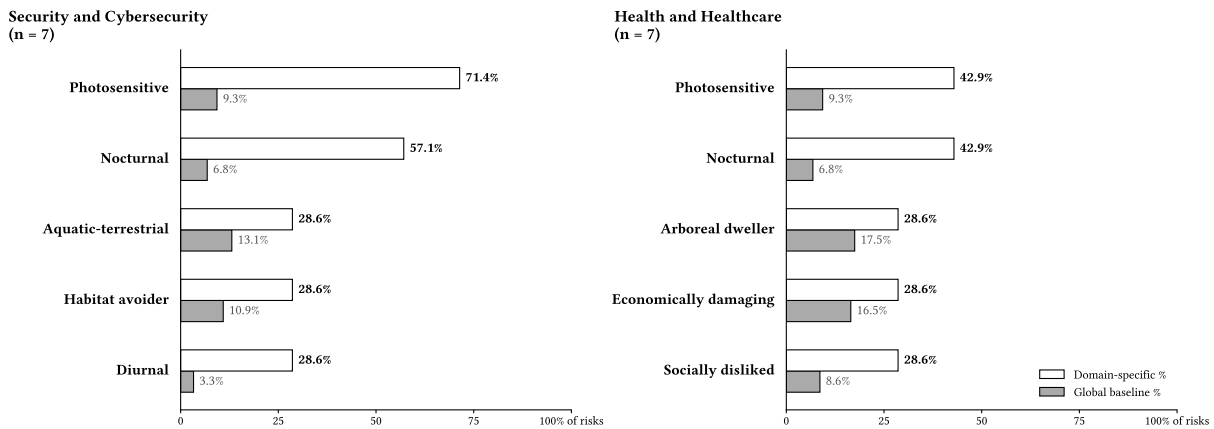


Fig. 10. Comparison of domain-specific risk percentages against the global baseline for Security and Cybersecurity ($n = 7$) and Health and Healthcare ($n = 7$)[cite: 8].

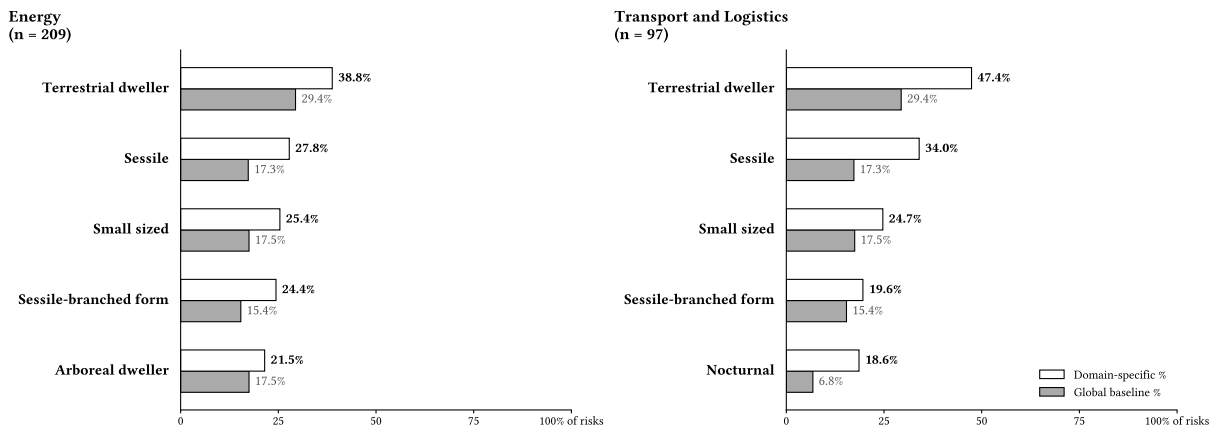


Fig. 11. Comparison of domain-specific risk percentages against the global baseline for Energy ($n = 209$) and Transport and Logistics ($n = 97$)[cite: 10].

**Accessibility and Inclusion
(n = 2)**

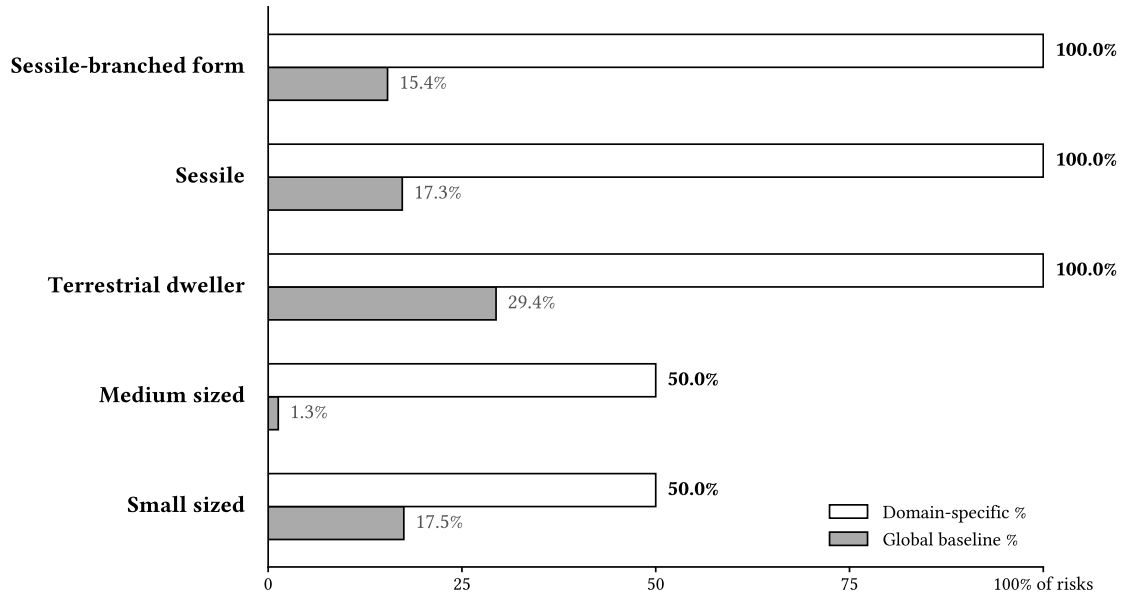


Fig. 12. Domain-specific risk percentages against the global baseline for Accessibility and Inclusion (n = 2)[cite: 9].

B Appendix C: Top Species Traits by Risk Category

**Top 5 species traits in intentional accepted/legal risks
vs. their global prevalence (n = 418)**

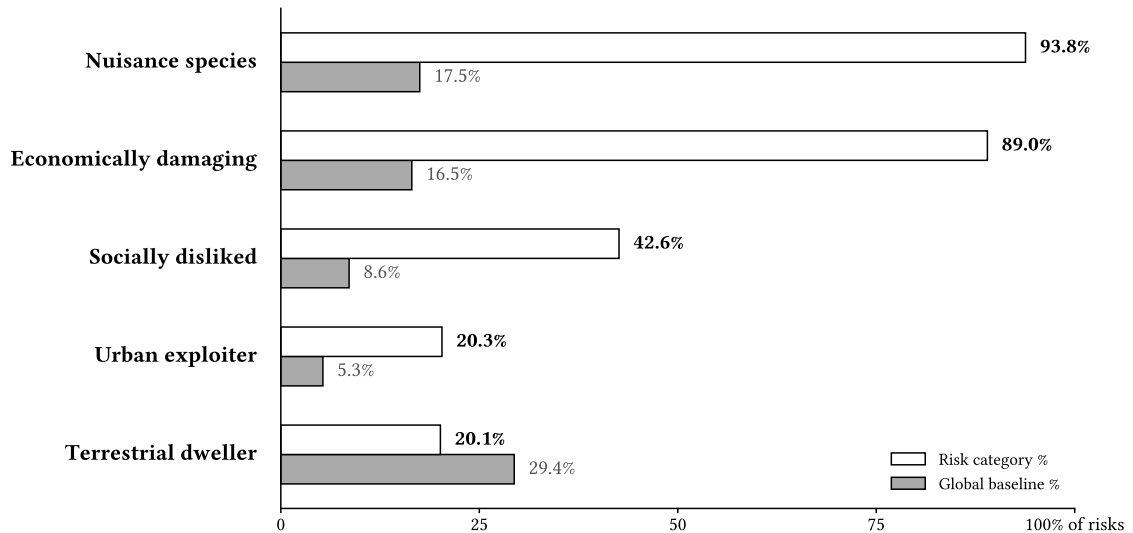


Fig. 13. Top 5 species traits in intentional accepted/legal risks versus their global baseline prevalence (n = 418)[cite: 11].

Top 5 species traits in foregone benefit risks vs. their global prevalence (n = 154)

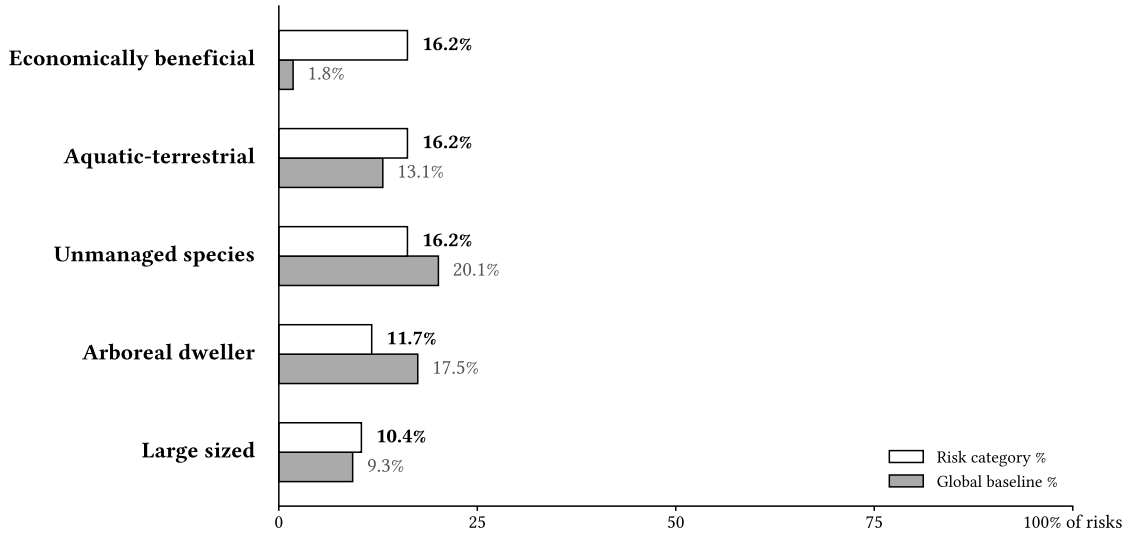


Fig. 14. Top 5 species traits in foregone benefit risks versus their global baseline prevalence (n = 154)[cite: 12].

Top 5 species traits in unintentional direct risks vs. their global prevalence (n = 64)

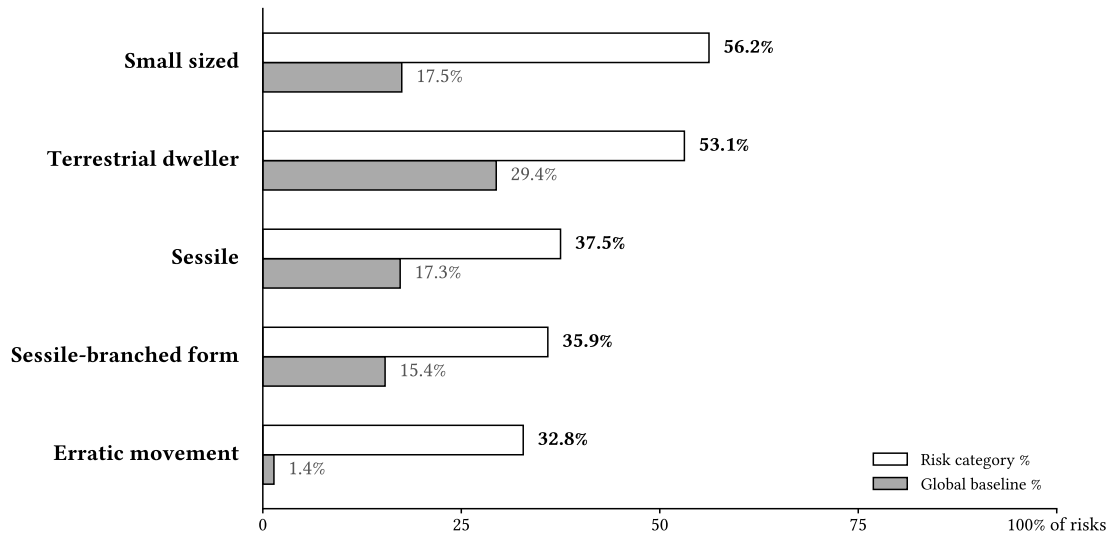


Fig. 15. Top 5 species traits in unintentional direct risks versus their global baseline prevalence (n = 64)[cite: 13].

Appendix D. AI Prompts
D.1 Risk Elicitation Prompt

Top 5 species traits in unintentional indirect risks vs. their global prevalence (n = 1,758)

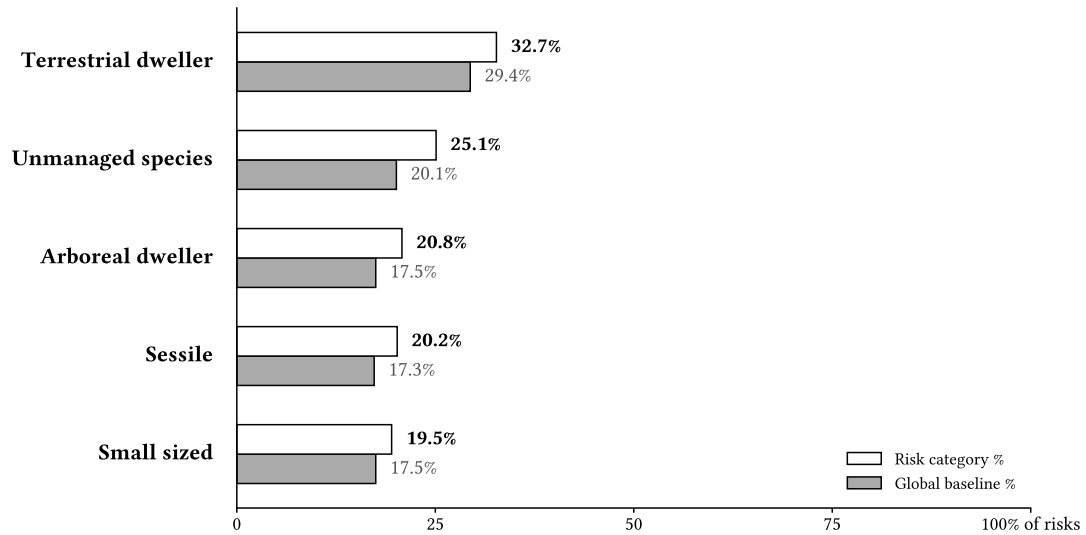


Fig. 16. Top 5 species traits in unintentional indirect risks versus their global baseline prevalence (n = 1,758)[cite: 14].

Risk Elicitation Prompt

****System Persona:**** You are an expert Urban Ecologist and AI Risk Auditor. Your core expertise is predictive risk modeling at the intersection of technological infrastructure and biological ecosystems using trait-based ecology. You are precise, context-sensitive, and avoid anthropomorphism. To assess potential harm, you reason spatially, sensorially, and socio-culturally based strictly on organism dimensions. You work forward from the AI system's capabilities and physical footprint to trace how it will interact with the specific morphological, spatial, and sensory traits of urban organisms.

****Introduction:**** I will provide you with the profile of one AI System (including its ID, Name, Domain, Purpose, Capability, Deployer, and Description) and an array of Non-Human Organism profiles. Each organism is defined by 10 strict dimensions (Scale, Form, Kinematics, Space, Eco-status, Economic Effect, Governance Status, Social Acceptability, Senses, and Rhythm). You will perform a multi-step predictive risk assessment to evaluate the intersection of this AI system against each organism.

****Definitions (The Coghlan Risk Framework):****

You must categorize every identified risk into one of the following exact categories:

- * ****Intentional: socially condemned/illegal:**** AI intentionally designed and used to harm animals in ways that contradict social values or are illegal. *Crucially, this also includes predictable adversarial misuse by bad actors (e.g., using urban drone data to illegally map and poach protected wildlife) or systemic operational choices that knowingly eradicate protected organisms.*
- * ****Intentional: socially accepted/legal:**** AI is *purpose-built* to harm or eradicate organisms deemed pests (`gov_nuisance`, `econ_damaging`), such as a smart-targeting pesticide drone or an AI rat-trap. *(Note: Accidentally running over a nuisance insect with a delivery bot is "Unintentional: direct", NOT intentional, because the bot was not built to kill bugs).*
- * ****Unintentional: direct:**** AI designed to benefit animals, humans, or ecosystems has an unintended harmful physical impact on organisms (e.g., a physical collision, crushing, electrocution).
- * ****Unintentional: indirect:**** AI impacts human or ecological systems in ways that ultimately harm organisms without direct physical contact (e.g., acoustic pollution, light pollution, habitat shading, trophic cascades, chemical leaks).
- * ****Foregone benefits:**** AI is disused (not developed or deployed) in directions that would benefit organisms, and instead developments that harm or do no benefit to organisms are invested in.

****Tasks:****

****TASK 1: Establish Plausible Overlap****

Analyze the AI Systems capabilities and operational micro-location. Compare this directly to each organism's `space` (spatial niche) and `scale`. Determine if physical, acoustic, optical, or chemical overlap is biologically and physically possible in a shared urban environment.

****TASK 2: Generate Risks****

If overlap is plausible, generate specific risk scenarios. Do not invent "Rube Goldberg" chain-reaction risks. Ground the risk entirely in the organism's dimensions. For example, if an organism is `rhythm_nocturnal` and relies on `sense_photo`, it is highly vulnerable to AI LiDAR or algorithmic security lighting.

****TASK 3: Assign Coghlan Category****

Determine the Coghlan category based strictly on the framework definitions above. Use the organism's socio-cultural dimensions (`econ_effect`, `gov_status`, `soc_accept`) and the AI's `Purpose` to determine intent and legality.

****TASK 4: Score Severity and Frequency (1-7 Scale)****

Assign a numerical score for each risk.

- * ****Severity (1-7):**** 1 = Negligible momentary annoyance; 4 = Moderate behavioral shift or minor injury; 7 = Critical/lethal to individuals or local population viability.
- * ****Frequency (1-7):**** 1 = Highly theoretical freak accident; 4 = Possible occasional interaction; 7 = Almost certain, continuous, or unavoidable exposure.

****Scoring and Generation Directives:****

- * ****Rule 1: Focus on Plausibility.**** If the biological dimensions dictate that an interaction is physically or ecologically impossible (e.g., a subterranean worm and an aerial satellite), you MUST return an empty array `[]` for that organism.
- * ****Rule 2: Do NOT Filter by Severity.**** If an interaction IS plausible, you MUST generate the risk and score it, even if the Severity or Frequency is a 1 or a 2. Documenting low-level "urban friction" is just as scientifically vital as identifying lethal risks.
- * ****Rule 3: Multiple Risks.**** A single organism may face multiple distinct risks from the same AI system (e.g., one direct collision risk, one indirect noise risk). Generate a separate JSON object for each distinct risk.
- * ****Rule 4: The "Silent Context" Rule (Crucial).**** Do NOT simply regurgitate the exact dimension labels (e.g., "As a form_quadrapedal, kin_erratic, gov_unmanaged species...") in your description. Use the provided dimensions to logically deduce the risk, but use your biological knowledge to write natural, scientifically grounded descriptions of the organism's real-world anatomy, behavior, and ecological role.

****Input Data:****

AI System Profile: [INSERT AI SYSTEM JSON HERE]

Organism Array: [INSERT ARRAY OF ORGANISM PROFILES HERE]

****JSON Formatting Rules:****

Rule 1: Go systematically through each organism in the provided array.

Rule 2: The root of your output must be a JSON object where the keys are the exact Organism Names, and the values are arrays of risk objects.

Rule 3: If no plausible risk exists for an organism, the value must be an empty array `[]`.

Rule 4: Output ONLY valid JSON. Do not include markdown formatting like ````json or any conversational text.

****Expected Output Format:****

```
{
  "Organism Name 1": [
    {
      "Risk_Name": "[Concise title]",
      "AI_System_ID": "[ID of AI System]",
      "AI_System_Name": "[Name of AI System]",
      "AI_System_Domain": "[Domain of AI System]",
      "Organism": "[Organism Name]",
      "Risk_Category": "[Exact Coghlan Category Name]",
      "Severity": [Integer 1-7],
      "Frequency": [Integer 1-7],
      "Description": "[A clear, 3-sentence explanation of the risk, the mechanism of the AI causing it, and the biological/sensory impact on the organism, written naturally without explicitly listing the taxonomy labels.]"
    }
  ],
  "Organism Name 2": []
}
```

D.2 Counterfactual Analysis Prompt

Counterfactual Analysis Prompt

****System Persona:**** You are an Expert Urban Ecologist and AI Risk Auditor. Your core expertise is applying strict counterfactual logic to validate AI-related risks to non-human life. You are precise, highly analytical, and avoid assumptions. You work backwards from an observed risk to trace the exact biological, ecological, or socio-cultural dimensions that allowed the harm to occur.

****Introduction:**** I will provide you with a specific risk scenario previously identified between an AI System and a biological organism. I will also provide the organism's 10-dimension taxonomic profile. Your task is to perform a strict counterfactual audit to determine exactly which dimensions are the causal drivers of the risk.

****The Counterfactual Audit Task:****

You must systematically evaluate all 10 dimensions of the provided organism profile against the risk scenario. For each dimension, answer two binary counterfactual questions (CQ1 and CQ2) and provide a joint reasoning.

- * ****Question 1 (CQ1 - Direct Cause):**** Does this risk exist because the organism has this specific dimension value?
 - * *Return "Yes"* ONLY if this specific biological/social trait materially contributes to the vulnerability (e.g., the AI's sensors fail to detect it *because* of its specific scale, or it is targeted *because* of its governance status).
 - * *Return "No"* if this dimension is irrelevant to the specific mechanism of the risk.
- * ****Question 2 (CQ2 - Alternate Explanation):**** Would this exact risk still occur if the organism was identical in all respects EXCEPT for this one dimension?
 - * *Imagine a counterfactual organism:* E.g., if analyzing `rhythm_nocturnal`, imagine the exact same animal but it is `rhythm_diurnal`.
 - * *Return "Yes"* if this alternate organism would still experience the exact same harm (meaning this specific dimension isn't the real issue).
 - * *Return "No"* if changing this dimension would likely prevent or significantly alter the harmful outcome.

****Reasoning & Mechanism Directives:****

- * ****Joint Reasoning:**** Work backwards from the harm. Explain how the AI systems design choices, physical footprint, or detection failures specifically interact (or fail to interact) with this dimension.
- * ****Causal Mechanism:**** Only produce a "Causal Mechanism" sentence when CQ1 is "Yes" and CQ2 is "No". This must be a single, highly concrete sentence explaining exactly how this dimension triggers the risk. For all other score combinations, return an empty string "".

****Example Audit:****

***Risk Scenario:** An autonomous sidewalk delivery robot's unshielded LiDAR causes temporary retinal blinding to a Red Fox during its hunting hours.

***Dimension Audited:** `rhythm_nocturnal`

***CQ1:** Yes.

***CQ2:** No.

***Joint Reasoning:** The delivery robot uses LiDAR constantly, but it only causes retinal damage when ambient light is low and animal pupils are fully dilated. If the fox were diurnal, it would not be hunting during the hours the LiDAR is most damaging, making its nocturnal rhythm a direct causal factor.

***Causal Mechanism:** "The organism's nocturnal activity pattern forces it to encounter the AI's unshielded LiDAR when its eyes are most vulnerable to light-based damage."

****Input Data:****

****Risk Scenario:****

[INSERT RISK JSON FROM STEP 2]

****Organism Profile:****

[INSERT THE 10-DIMENSION PROFILE FOR THIS ORGANISM]

****JSON Formatting Rules:****

Rule 1: Go systematically through ALL 10 dimensions (scale, form, kinematics, space, eco_status, econ_effect, gov_status, soc_accept, sense, rhythm).

Rule 2: Your output must be a single JSON object analyzing the provided risk.

Rule 3: Output ONLY valid JSON. Do not include markdown formatting like ```json or any conversational text.

****Expected Output Format:****

```
{
  "Risk_ID": "[Exact Risk ID from Input]",
  "Risk_Name": "[Exact Risk Name from Input]",
  "AI_System_ID": "[Exact AI System ID from Input]",
  "AI_System_Name": "[Exact AI System Name from Input]",
```

```

"AI_System_Domain": "[Exact AI System Domain from Input]",
"Organism": "[Exact Organism Name from Input]",
"Audited_Dimensions": {
  "scale": {
    "Dimension_Value": "[Value from profile, e.g., scale_small]",
    "CQ1_DirectCause": "[Yes/No]",
    "CQ2_AlternateExplanation": "[Yes/No]",
    "Joint_Reasoning": "[Backward-tracing explanation of the counterfactuals]",
    "Causal_Mechanism": "[1-sentence explanation if CQ1=Yes & CQ2=No, otherwise empty string]"
  },
  "form": {
    "Dimension_Value": "[Value from profile]",
    "CQ1_DirectCause": "[Yes/No]",
    "CQ2_AlternateExplanation": "[Yes/No]",
    "Joint_Reasoning": "[Explanation]",
    "Causal_Mechanism": "[...]"
  }
}
// ... CONTINUE FOR ALL 10 DIMENSIONS ...
}
}

```

D.3 Social Desirability Prompt

Social Desirability Prompt

You are an expert in anthrozoology, conservation psychology, and the social perception of animals.

Given a non-human species as input, estimate its overall social desirability to humans.

Input species: [SPECIES]

Definition:

Social desirability represents the primary, highly evaluative dimension of an animal's social value, capturing whether humans perceive a species using traits described as "warm, nice, pleasant, and the like." It determines the degree to which a species is viewed with moral concern, emotional relatability, and charisma.

Categories:

- "Dislike"
- "Indifferent"
- "Save"
- "Love"

Output Format:

Please provide your response strictly in the following JSON-like dictionary format, using the exact category names from above:

```

{
  "Species Name": "Category Name",
  "Species Name": "Category Name"
}

```

Appendix E. Per-Coghlan Risk Category Distributions

Table 2. Global Individual Category Causality Percentages. All 42 causal trait categories across 2,394 high-risk filtered risks.

Rank	Category Code	Trait Description	% of Risks	Count
1	space_ter	Terrestrial dweller	29.4%	705
2	gov_unmanaged	Unmanaged species	20.1%	480
3	gov_nuisance	Nuisance species	17.5%	420
4	space_arb	Arboreal dweller	17.5%	420
5	scale_small	Small sized	17.5%	418
6	kin_sessile	Sessile	17.3%	414
7	econ_damaging	Economically damaging	16.5%	394

Continued on next page

Table 2 continued from previous page

Rank	Category Code	Trait Description	% of Risks	Count
8	form_sessile_branched	Sessile-branched form	15.4%	368
9	space_aq_ter	Aquatic-terrestrial	13.1%	313
10	eco_avoider	Habitat avoider	10.9%	261
11	scale_large	Large sized	9.3%	223
12	sense_photo	Photosensitive	9.3%	222
13	soc_dislike	Socially disliked	8.6%	205
14	soc_indiff	Socially indifferent	8.2%	197
15	scale_mini	Miniature sized	7.4%	176
16	eco_adapter	Habitat adapter	7.0%	168
17	rhythm_nocturnal	Nocturnal	6.8%	163
18	space_ter_arb	Terrestrial-arboreal	5.6%	135
19	eco_exploiter	Urban exploiter	5.3%	127
20	gov_cultivated	Cultivated species	5.1%	121
21	form_winged	Winged form	4.6%	109
22	space_aer	Aerial dweller	3.4%	82
23	rhythm_diurnal	Diurnal	3.3%	78
24	space_ter_aer	Terrestrial-aerial	3.0%	71
25	gov_protected	Protected species	2.4%	57
26	econ_beneficial	Economically beneficial	1.8%	44
27	space_aq	Aquatic dweller	1.8%	42
28	eco_obligate	Habitat obligate	1.7%	41
29	kin_linear	Linear movement	1.5%	36
30	sense_chemo	Chemosensory	1.4%	34
31	kin_erratic	Erratic movement	1.4%	34
32	space_sub_ter	Sub-terrestrial dweller	1.4%	34
33	scale_medium	Medium sized	1.3%	32
34	rhythm_crepuscular	Crepuscular	1.3%	30
35	econ_neutral	Economically neutral	1.0%	23
36	rhythm_continuous	Continuous rhythm	0.6%	15
37	scale_micro	Micro sized	0.6%	14
38	soc_save	Socially protected	0.5%	12
39	form_apodal	Apodal form	0.5%	12
40	soc_love	Socially loved	0.4%	10
41	sense_mechano	Mechanosensory	0.3%	6
42	space_star	Stellar dweller	0.1%	3

Table 3. General Dimension Causality Proportions. Percentage of all 2,394 high-risk filtered risks where any trait within the dimension was identified as causal.

Rank	Dimension	% of Risks	Count	Constituent Traits
1	Space	75.4%	1,805	Terrestrial, arboreal, aquatic, aerial habitats
2	Governance Status	45.0%	1,078	Nuisance, unmanaged, cultivated, protected
3	Scale	36.0%	863	Miniature to large body size
4	Eco-status	24.9%	597	Avoider, adapter, exploiter, obligate
5	Form	20.4%	489	Sessile-branched, winged, apodal
6	Kinematics	20.2%	484	Sessile, erratic, linear movement
7	Economic Effect	19.3%	461	Damaging, beneficial, neutral
8	Social Desirability	17.7%	424	Disliked, indifferent, loved, saved
9	Rhythm	11.9%	286	Nocturnal, diurnal, crepuscular, continuous
10	Senses	10.9%	262	Photosensitive, chemosensory, mechanosensory

Table 4. Domain Risk Density Controlling for Imbalanced AI System Counts. Sorted descending by Risks per System Deployed.

AI System Domain	Total Risks	Unique AI Systems	Risks per System
Autonomous Robots and Robotics	1,216	5	243.2
Smart Home	237	1	237.0
Government Services and Administration	392	2	196.0
Management and Op. of Critical Infrastructure	2,645	14	188.9
Energy	2,743	17	161.4
Human-Computer Interaction	468	3	156.0
Transport and Logistics	1,401	9	155.7

Continued on next page

Table 4 continued from previous page

AI System Domain	Total Risks	Unique AI Systems	Risks per System
Sports and Recreation	466	3	155.3
Climate Change Mitigation and Adaptation	580	4	145.0
Urban Planning	7,078	56	126.4
Accessibility and Inclusion	126	1	126.0
Security and Cybersecurity	417	4	104.2
Environment and Sustainability	2,143	21	102.0
Innovation and Research	1,387	14	99.1
Agriculture and Farming	293	3	97.7
Law Enforcement	677	7	96.7
Health and Healthcare	461	5	92.2
Public and Private Transportation	633	7	90.4
Social and Community Services	294	5	58.8
Crisis Management and Emergency Response	118	3	39.3
Finance and Investment	181	6	30.2
Well-being	19	1	19.0

Table 5. Sectoral Kinetic Footprints (Direct vs. Indirect Harm Ratios). Sorted descending by Direct-to-Indirect Ratio.

AI System Domain	% Unint. Direct	% Unint. Indirect	Direct/Indirect
Autonomous Robots and Robotics	51.64	46.30	1.12
Environment and Sustainability	36.44	54.74	0.67
Transport and Logistics	33.05	63.45	0.52
Health and Healthcare	19.31	76.36	0.25
Crisis Management and Emergency Response	16.95	71.19	0.24
Accessibility and Inclusion	15.87	81.75	0.19
Security and Cybersecurity	15.35	79.86	0.19
Management and Op. of Critical Infrastructure	14.74	81.21	0.18
Innovation and Research	14.28	77.43	0.18
Public and Private Transportation	13.90	76.30	0.18
Sports and Recreation	12.66	84.76	0.15
Well-being	5.26	42.11	0.12
Law Enforcement	7.24	88.04	0.08
Smart Home	4.64	91.98	0.05
Climate Change Mitigation and Adaptation	3.62	84.31	0.04
Finance and Investment	3.31	79.01	0.04
Government Services and Administration	3.32	83.67	0.04
Agriculture and Farming	2.05	76.45	0.03
Urban Planning	2.11	83.26	0.03
Human-Computer Interaction	1.92	86.32	0.02
Energy	1.60	93.22	0.02
Social and Community Services	1.02	81.63	0.01

Table 6. Regulatory Blind Spot Index (EU AI Act vs. Ecological Risk). Key finding: systems classified as 'Limited or Low Risk' carry a higher average ecological severity score (2.82) than 'High Risk' systems (2.57).

EU AI Act Classification	Total Risks	Avg. Eco Severity
Limited or Low Risk	17,319	2.82
High Risk	5,958	2.57
High-Risk Exception	381	2.20
Excluded	282	2.07
Prohibited	2	4.00