

Empirical Self-Improvement in Normative Reasoning

Autonomous Orchestration and a Computable Theory of Normative Choice

JudgeAI Research

Revised 18 September 2026

Abstract

JudgeAI combines a computable theory of normative choice with an autonomous AI orchestrator that develops its empirical grounds. The computation identifies how consequences for affected actors enter the comparison of regulatory worlds. Its trace exposes the dependencies through which further evidence can alter that comparison. The orchestrator follows these dependencies through the ontology, performs targeted research, analyses the applicability and structure of the evidence, develops additional alternatives and submits revised worlds to the same normative engine. Each evaluation therefore supplies both a decision and a basis for the next investigation. We examine this process through completed work on Amsterdam tourism and private-rental energy improvements in England and Wales, with a financial-services study illustrating decision-focused empirical inquiry in a further application. In Amsterdam, empirical analysis participated in selecting the governing regulatory course and subsequently deepened the financial basis of the package. In the British case, six linked comparisons developed a broad policy question into connected provisions for finance, staged work, occupation and effective repair. These cases demonstrate how traceability functions within autonomous reasoning: it turns a normative decision into an object of continued empirical investigation. Self-improvement consists in an increasingly informed representation of reality, more precisely specified alternatives and a regulatory design whose interdependent conditions are evaluated under a stable normative foundation.

Keywords: computational legal theory; normative reasoning; traceability; empirical inquiry; AI orchestration; uncertainty; JudgeAI.

1 Autonomous empirical improvement in normative reasoning

The central property examined in this article is JudgeAI's capacity to improve a normative decision through continued autonomous investigation of reality. The AI orchestrator works with an executable theory of normative choice. It studies the trace of a calculation, identifies the empirical distinctions that matter to that choice, investigates them through the ontology and returns a more developed representation of the situation to the engine. The new calculation supplies another decision and another basis for inquiry. Research and normative computation thus form a continuing process in which each develops the work of the other.

This interaction addresses a distinctive condition of regulatory design. A proposed law must be considered before its consequences have fully materialised. Its effects depend on how businesses adapt, how households experience costs, how institutions deliver services and how several provisions operate together. An immediately available, generally accepted benchmark does not settle the quality of every such decision. Implementation later supplies evidence about what happened; during design, the system needs feedback that connects present investigation to the grounds of normative choice.

JudgeAI obtains that feedback from its computational legal theory. The theory supplies the substantive basis on which consequences for affected actors are compared. The ontology makes those actors,

relationships, actions and consequences available as explicit objects of investigation. The mathematical apparatus evaluates the represented worlds and records how their empirical content enters the result. This makes further work specific: the orchestrator can investigate a particular financing relationship, exposure pathway, affected population or condition of execution because its place in the normative comparison is known.

The theoretical foundation is developed in the published JudgeAI study *Law as Computation: An End-to-End System for Normative Decision-Making* [1]. This article examines the autonomous process this foundation enables. Its subject is the reciprocal relationship between empirical inquiry and computable normative choice, demonstrated through the decisions and regulatory texts produced in completed cases. The contribution is a working account of how a traceable normative foundation becomes an instrument for sustained improvement of the decision itself.

2 A computational trace as a working object

2.1 What the normative foundation contributes

JudgeAI’s legal theory supplies the substantive basis of normative comparison. It represents the effects of arrangements on actors and their possibilities of viable action, rather than treating an LLM’s preference between written answers as the final decision rule. The ontology gives the theory an addressable subject matter: actors, roles, relationships, actions, contexts, and consequences. Empirical calibration connects these objects to evidence about the world. The engine applies the specified comparison to the resulting representations [1].

This foundation is essential to the interaction examined here. It establishes what an empirical change is a change *to*, and how that change can matter normatively. An observation about business liquidity, for example, must have a defined relationship to an actor’s exposure and to the consequences of a proposed obligation. Its role cannot be established merely by including the observation in the same document as the recommendation.

The theory determines the normative grounds of comparison; inquiry investigates the reality to which they apply. This division gives empirical improvement a stable meaning. When the orchestrator establishes a different distribution of burdens, discovers an executable alternative or refines a causal relationship, the engine evaluates the resulting change under the same substantive rule. The connection between learning about reality and making a normative choice is part of the system’s computational structure.

2.2 The evaluated state and its trace

For the purposes of this article, an evaluated state at stage t is represented by

$$S_t = (O_t, \mathcal{E}_t, M_t, \mathcal{W}_t, \mathcal{U}_t), \quad (1)$$

where $\mathcal{E}_t$ is the evidence record, M_t the model of consequences and its calibration, $\mathcal{W}_t$ the represented regulatory alternatives, and $\mathcal{U}_t$ the treatment of empirical uncertainty. These objects are related: an empirical record refers to a population, action or exposure, comparator, outcome, context and horizon represented in the ontology.

Let $\mathcal{T}^{(v)}$ denote a specified version of the normative foundation and let $F_{\mathcal{T}^{(v)}}$ be its computational implementation. We write

$$(D_t, \tau_t) = F_{\mathcal{T}^{(v)}}(S_t), \quad (2)$$

where D_t contains the normative comparison and its selected alternatives, and τ_t records the executed dependencies and operations. The notation accommodates a set of equally ranked alternatives; a single scalar score is not assumed as a universal form of normative theory. Equation (2) describes the

research–computation interface. The published theoretical account develops the normative foundation; the following sections explain how its computational trace supports the autonomous inquiry studied here.

A trace suitable for this interaction connects the evidence used, the interpretation and calibration applied, the represented consequences and their contribution to the comparison. Every dependency used by the executed normative calculation has an explicit representation that can be followed to its recorded inputs or assumptions. The status of each input travels with it. The orchestrator can therefore continue from a particular dependency, examine its evidence and inspect the consequences of a revision throughout the calculation.

The resulting object is usable by an agent. The orchestrator can ask which actor, transition, estimate or applicability condition contributes to a result; what evidence supports it; and what changes when an empirical assumption is varied. A narrative explanation can present this information, but the operational basis is the record of execution.

2.3 The trace identifies the next empirical question

Two evaluations can produce the same recommendation through different empirical grounds. A financing proposal might be preferred because a burden is dispersed across actors with available resources, or because the representation has not yet distinguished the people carrying it. A printed recommendation alone leaves these states indistinguishable. Their traces identify different next questions: one concerns the stability of a supported incidence model; the other concerns the joint financial information needed to develop that model.

The trace supplies the orchestrator with the location and meaning of that difference. It connects an actor’s financial state to the relevant tax obligation, ownership relation, cash timing and empirical source. Additional research can then change the appropriate part of the represented world and be checked through another calculation. In this architecture, traceability is the working memory and feedback interface of empirical normative reasoning.

3 The autonomous inquiry–computation cycle

The orchestrator carries out the research operations required by the normative pipeline. It formulates successive task requests, follows the ontology-defined research targets, acquires and analyses evidence, supplies the resulting empirical representation and reads the engine’s comparison. The pipeline determines the semantic, evidential and computational requirements; the normative choice is performed by the engine. The next request develops a dependency revealed by the preceding work.

Let an inquiry q_t be formed from the task state, the decision and its trace. Its result r_t updates the empirical representation:

$$q_t = \Pi(\mathcal{T}^{(v)}, \mathcal{S}_t, D_t, \tau_t), \quad (3)$$

$$\mathcal{S}_{t+1} = \text{Update}(\mathcal{S}_t, r_t), \quad r_t = \text{Investigate}(q_t). \quad (4)$$

The updated state is evaluated by Equation (2). The orchestration function Π is realised by the autonomous AI agent operating through the established pipeline. It uses the meaning of ontology relations and the executed dependencies to formulate a relevant next request. A question about a tenant’s available money, for example, leads to evidence about actual energy payments, attributable rent and disruption costs over a specified period. A question about a landlord’s capacity to finance work leads to available cash, binding funding, liabilities and payment timing.

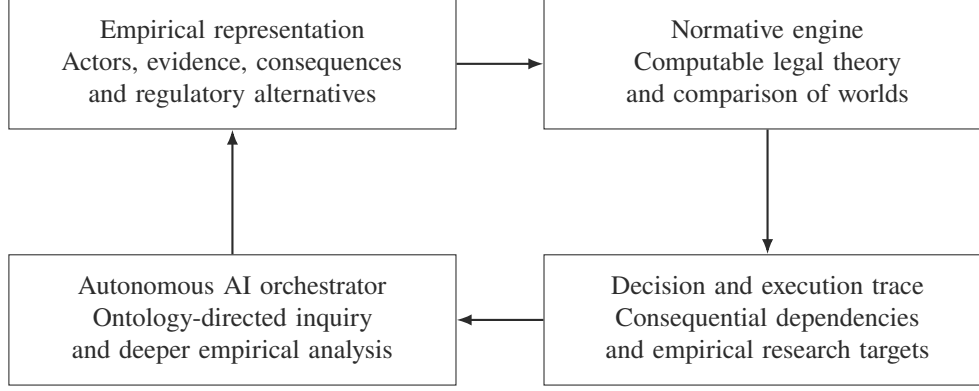


Figure 1: Autonomous empirical self-improvement. Computation supplies the grounds of the next inquiry; inquiry develops the worlds submitted to the next computation. The normative criterion remains fixed through the empirical revision.

3.1 The ontology gives research a computational destination

The ontology provides more than a vocabulary for finding documents. It specifies the relationship that a source must help explain. A research target identifies an affected actor, the relevant role, the intervention or exposure, a consequence, a comparator and a horizon. The orchestrator can move from a calculated effect to the facts needed to assess it, and from a new finding back to the part of the calculation to which it applies.

This makes inquiry both more targeted and deeper. In the British housing case, the relationship between the tenant’s living environment and health led to research on moisture, ventilation and installation defects. The payment-obligation relationship led to net household resources after bills, rent and disruption. The owner-resource relationship led to the distinction between a property valuation and cash available when an invoice falls due. The same dwelling thus generates several precise empirical questions, each with its own outcome and evidence requirements.

The source’s meaning remains attached to the research target. A metered reduction in gas consumption concerns energy use over a stated interval; a medical study may concern consultation frequency within a particular population; a funding programme establishes eligibility and award conditions. The orchestrator investigates how these findings support the relevant consequence instead of treating their numerical values as interchangeable. This semantic precision is a direct contribution of the ontology to the quality of subsequent computation.

3.2 Each result develops the next task

Empirical development can refine an existing quantity, distinguish a population, reconstruct a dependence between variables, specify a causal mechanism or add an alternative made feasible by newly examined conditions. A funding source, for example, can create an executable improvement option that was absent from a comparison framed only around compulsory expenditure. The new option receives the same normative evaluation as the others.

The trace also follows the transmission of a revision. The orchestrator checks that the new evidence is represented in the accepted inputs, that the intended actors and consequences are retained, and that the revised state reaches the mathematical comparison. This joins research, calibration and evaluation into a single inspectable process. It supplies a computational check on whether deeper empirical work has actually developed the decision.

The object that develops through this cycle is the system’s task-specific knowledge, representation of alternatives and normative solution. Each evaluated state retains its evidence and conditions of applicability. That state becomes the starting point for the next inquiry, so further work can extend

the same solution rather than regenerate an unrelated answer. A stable decision can acquire a richer empirical justification; a new consequence or alternative can produce a different decision. Both forms of development are visible through the trace.

4 Empirical depth and the role of uncertainty

4.1 From unavailable evidence to a researchable question

Traceability turns an empirical dependency into a researchable object. Where a relevant quantity requires further evidence, the system retains its actor, context, outcome and place in the comparison. The next inquiry can then seek the particular measurement or relationship needed to develop that consequence.

In JudgeAI’s calibration framework, research targets are tied to actor-context transitions. They identify the affected population, exposure, comparator, outcome and horizon. The calibration layer distinguishes evidence directly supporting a target from evidence requiring contextual transfer. It also tracks study and cohort identity, dependence among observations, and the requirements for identifying a parameter or contrast.

Tensors provide a structured mathematical representation of empirical quantities in the normative model. Their coordinates and relations make it possible to locate the combinations supported by measurements and identify the additional measurements needed for a target. In a locally linear calibration, for example, observations of certain combinations need not identify every target contrast. Additional publications about the same observed direction do not automatically identify a missing direction. The recorded target can therefore specify the additional kind of measurement required.

The implementation turns these distinctions into research requests. A request carries the target, existing sources, required measurement and unresolved conditions. Proposed precision gains are conditional on the stated measurement design and dependence assumptions. The mathematical diagnosis supplies a question for investigation; empirical evidence supplies its answer. This is a concrete route by which computation changes the next activity of the orchestrator.

4.2 More depth can mean a different empirical structure

The depth of inquiry is not measured by the number of documents retrieved. For regulatory design, moving from separate means to a joint distribution can be more consequential than adding another estimate of an already known average. An annual charge linked to property value, for example, interacts with the owner’s cash income, financing obligations, rental position and compliance costs. The mean of each variable does not establish which businesses bear the combined burden.

Empirical deepening can therefore involve disaggregation, matching, reconstruction of dependence, examination of causal mechanisms, or assessment of transfer to the relevant setting. Each produces a more differentiated account of who is affected and how. The evidential status of the reconstruction must be preserved: an observed joint sample and a calibrated forecast can both enter analysis, but they supply different kinds of knowledge.

This distinction is productive for orchestration. A forecast identifies a dependency worth evaluating now and a measurement worth seeking later. It makes a hypothesis testable within the represented worlds without presenting that hypothesis as a directly observed fact. The orchestrator can subsequently replace, refine or contest it using further evidence.

4.3 Robustness makes the consequences of uncertainty visible

For a represented set of empirically plausible scenarios $\mathcal{U}_t$, the engine can compare the alternatives under each scenario. Let $D_t(u)$ denote the resulting selection at $u \in \mathcal{U}_t$. A change in $D_t(u)$ identifies a

decision-relevant contrast; further calculation can locate a transition within the specified empirical domain. Stable selections identify the evaluated scenarios in which the preferred alternative is preserved; transitions identify where further empirical precision can affect the choice.

This analysis separates two questions: how uncertain an input is, and how consequential that uncertainty is for normative choice. An imprecise parameter may have little influence on the preference between alternatives. A much narrower uncertainty may matter greatly if it straddles a decision boundary. Preliminary calculation can therefore help the orchestrator formulate a more useful research question than a general request for additional information.

The same logic applies to joint assumptions and model structure. Correlated costs can change burdens in a way that isolated parameter changes do not expose. An alternative causal model may change which consequences are expected at all. The examined uncertainty set should identify these distinctions so that a statement of stability has a defined empirical scope.

Robustness is a characteristic of the evaluated decision. Its normative role follows the decision procedure in use. If that procedure evaluates a specified risk measure, that measure is computed. The orchestrator uses the resulting sensitivity information to direct empirical inquiry under that same procedure.

5 Completed JudgeAI studies

The following completed studies show autonomous orchestration working through successive empirical and computational tasks. Amsterdam and the British rental-energy case use JudgeAI’s native normative engine: the former reaches the governing policy course and financing incidence; the latter follows the development of an integrated legislative design. A financial-services study then illustrates how ontology-linked inquiry and sensitivity analysis operate in a task-specific legal and economic configuration kernel. In each case, the relevant evidence, represented world and result can be connected through the execution record.

5.1 Amsterdam: empirical analysis reaches the governing policy choice

The principal substantive result of the Amsterdam study appeared at the first policy decision. The task was to develop a concrete, coherent regulatory package for tourism on the basis of independently researched current circumstances. It left the governing policy direction open. The supplied bare GPT-6 package retained a citywide objective of reducing overnight stays to 20 million, with a prescribed trajectory of 23 million in 2027, 22 million in 2028, 21 million in 2029 and 20 million from 2030. JudgeAI selected a different governing approach: regulation of specific local impacts and service capacity.

This difference arose at the level at which the package’s purpose was operationalised. JudgeAI made the overall regulatory course a question for the native engine. Five alternatives entered the recorded comparison: expanding permitted tourism capacity; contracting it; regulating specific impacts; retaining the existing architecture; and ending organised commercial tourism. Alternatives requiring legislative amendment remained available for evaluation. The first normative task therefore concerned which general arrangement should govern the subsequent package, including whether an aggregate volume objective should be retained.

Empirical inquiry informed the consequences attributed to these competing arrangements before the engine selected one. The initial evidence record drew on local tourism pressure, visitor forecasts, hospitality business conditions and employment. The 2025 municipal neighbourhood study identified six neighbourhoods under pressure, while also documenting improvements in Grachtengordel-Zuid and Oosterparkbuurt. The observations distinguished tourism intensity from the conditions experienced in particular places [3]. The municipal forecast, meanwhile, anticipated between 25.0 and 29.4 million

overnight stays in 2028 [4]. These sources supplied different parts of the empirical account: the spatial distribution of pressure and the prospective scale of demand.

The consequence model connected that account to the affected actors. Expansion had to be assessed alongside disturbance in already pressured areas. Contraction had to be assessed alongside its implications for business activity, employment and access. National business evidence was used with an explicit distinction between sector-level performance and an individual owner's financial resources. The initial record also avoided treating reduced tourism capacity as an automatic conversion of premises into housing. Its consequence estimates were recorded as model-inferred ordinal mappings, with a common baseline and horizon across alternatives. Published observations and the inferences used to compare possible worlds consequently remained distinguishable.

The native comparison selected regulation of specific impacts without a procedural tie-break. The resulting legislation expressed that choice directly. Article 3 proposed abolishing the mandatory city-wide overnight-stay range. Article 1 allowed activity at a particular location to increase, remain stable or decrease according to the applicable conditions. Admission, operational obligations and implementation capacity were then developed within this selected architecture. These provisions appear in the published consolidated package of draft regulatory acts [2].

The empirical analysis thus participated in determining the regulatory course itself. The supplied GPT-6 package also used municipal evidence, including in its zoning provisions. The distinctive result here is that the relationship between evidence, represented consequences and the governing choice was made executable: the continuation of the existing course was evaluated as an alternative, and another course was selected. The comparison can be reopened through its recorded empirical dependencies. Further research can alter the represented effects on residents, operators, workers or visitors and enter another evaluation under the same normative foundation.

This is why the Amsterdam case is more than a contrast between two legislative texts. It shows the substantive reach of the interaction: empirical inquiry can enter the choice that organises the whole package, while the resulting computational representation identifies where deeper inquiry can develop that choice. The initial comparison determined the architecture of the package. Its subsequent development required investigating empirical dependencies within that architecture.

5.2 Amsterdam: investigating the incidence of a financing decision

The Amsterdam study concerned a coordinated package of tourism regulation. A financing question emerged within that package: whether to let an expiring annual EUR 10 million appropriation lapse and defer unfunded new duties, or to renew that appropriation through property taxation (OZB) or an additional tourist-tax burden. The alternatives could not be adequately compared by listing their statutory tax bases. Their consequences depended on who would bear the payments, how tourism demand would respond, and how the resulting burden would interact with the resources of businesses and their owners.

The traceable representation made this empirical question specific. A financing measure affected actors through an estimated reduction in available resources. To assess that effect, the analysis needed a relationship between the obligation and the actor's financial position. The relevant inquiry consequently extended to property values, bank balances, income, rent, debt and the costs already included in the operating account. A separate average for each variable would not identify their joint incidence.

The completed central calibration addressed this question by combining published Statistics Netherlands (CBS) information on entrepreneurial income and wealth composition. Wealth strata supplied a common grouping through which assets, mortgage debt and other debt could be related. Within strata, a bank-balance distribution was reconstructed from published means and medians. Business turnover was connected to entrepreneurial income through a published sector margin. Existing premises costs, business interest and compliance expenses were treated through their position in the underlying net business income, preventing them from being deducted a second time [5, 6].

The research also refined the consequence mechanism for the tourist-tax alternative. The analysis used the central Amsterdam APE estimate with a 75% tax pass-through assumption. The demand response was derived from the study's treatment of accommodation expenditure within the total cost of a trip, and estimated procurement savings were deducted from the lost revenue. This made the forecast sensitive to the relevant economic mechanism, rather than treating an accommodation-price change as an equal percentage change in the cost of the whole trip [7].

The resulting joint account was a statistically matched central forecast, with recorded assumptions, population mapping and imputed components. It entered the computation as a specified empirical hypothesis whose individual dependencies remained available for further investigation. The key development was the move from loosely connected financial quantities to a model of how the financing alternatives would affect the resources of the represented actors.

One diagnostic in the completed calibration, for the modelled tourism-business owners, was the expected ratio of additional personal liquidity need over five years to the initial bank balance, capped at one. It was approximately 8.5% under the OZB alternative and 21.7% under the tourist-tax alternative. These quantities are modelled cash-resource indicators, defined by the additional liquidity requirement relative to initial balances. They were mapped into the corresponding empirical inputs of the normative representation. The native engine compared the complete profiles in the supplied, versioned actor field, including the other consequences of the financing arrangements.

The recorded result selected the property-tax financing alternative with a non-tied numerical preference under the kernel comparison. The verification record identifies the existing geometry, tensor, causal-propagation and TI/minimax kernel as the mathematical authority. The autonomous orchestrator followed the evidence through its numerical representation, froze the accepted calibration before evaluation and checked all 84 primary input values against the native result.

The proposed financing provision retained its property-tax basis. What improved was the empirical ground on which that provision was preferred. The resulting bill connected the additional funding requirement to the relevant property-tax revenue forecast, preserved the distinctions between tax categories, and required the appropriation and rate changes to be implemented together [2]. The deeper investigation made the justification of that arrangement more differentiated and its conditions more explicit.

This episode is especially informative because improvement did not require a different headline recommendation. The represented dependence between financial resources and regulatory burden became more substantive; the engine could distinguish the alternatives on that basis. The interaction also produced a more focused future question: whether additional evidence about ownership, liquidity and joint financial exposure would preserve the comparison. The computation supplied the place where those facts mattered, and inquiry supplied a more developed account of them.

5.3 United Kingdom: autonomous development of a connected regulatory design

The British task concerned a comprehensive legislative response to a current regulatory problem. The orchestrator investigated energy improvements in privately rented homes in England and Wales and developed the proposal through six completed comparisons: 26 option-world evaluations and 46 within-comparison pairs. These counts include alternatives evaluated in more than one question. Each comparison used the existing ontology and normative criterion.

The empirical record combined several kinds of evidence. The government's 2026 impact assessment described forecast installations, owner costs and tenant savings for higher performance standards. The MEES evaluation supplied administrative and survey evidence about the existing regime. NEED supplied matched, metered evidence about energy use after particular measures. Welsh health research examined outcomes in its defined recipient population, while the ECO4/GBIS audit identified specific installation and safety mechanisms. Funding guidance supplied the conditions under which a project could actually obtain support [13, 14, 15, 16, 17, 18].

The computational comparisons represented these mechanisms in declared conditional worlds. Reference, favourable and adverse ordinal input states were 0.50, 0.35 and 0.65 respectively, with reliability 0.50 under the existing input contract; the defect-response question used 0.65 for the identified defective starting condition. Observed measurements and official forecasts supplied their own outcomes, populations and units, while these ordinal states encoded the specified comparison conditions. Unestimated channels were held common within the comparisons as an explicit scope condition. The scientific result is the sequence of empirical distinctions, executable alternatives and connected provisions obtained by developing and recomputing those worlds.

From a broad policy choice to its financing dependency. The first question compared retaining or removing the current floor, voluntary improvement and three higher mandatory configurations. The trace identified uncompensated landlord expenditure as the decisive burden separating the higher-standard group from the leading group under the supplied profiles. Retention, repeal and unspecified voluntary improvement had equal calculated consequences. The orchestrator followed the financing dependency into a more specific question: what changes when effective improvements have binding eligible funding, covered delivery costs and positive net tenant savings?

Warm Homes Local Grant guidance established concrete eligibility and contribution arrangements worth evaluating. The next comparison introduced a complete funded-delivery alternative. Its conditions included safe execution, actual finance, covered administration and continued provision for other recipients. The engine selected this alternative within the revised set. The inquiry had developed both the empirical description and the available choice: the comparison now included how the improvement could be paid for and delivered, alongside what performance was desired.

From a legal label to observable delivery. A further question compared a statutory process with equivalent delivery under existing powers and matched delivery alongside repeal. Funding, completed work, occupation and costs were held equal. The engine returned an exact tie. This supplied a precise research direction: the distinguishing evidence would concern changes in actual delivery, enforceability, delay, costs or participation. Equal realised consequences received equal treatment, focusing subsequent inquiry on what a legal arrangement would change in the represented world.

From whole-project funding to an independently useful stage. The next task addressed a funding gap. The orchestrator represented a safe, technically separable stage with its complete costs covered, then compared it with tenant-gap funding, deferral and uncompensated landlord-gap funding. The funded stage ranked first. The resulting legislative provision permits staging when the stage has an independently verified benefit, is safe without reliance on unfunded later work and preserves the practicability of that later work. A partial award thus became a precisely bounded delivery option rather than an assumed completed project.

From installation paperwork to effective repair. Ontology-linked inquiry followed the tenant's living-environment and health relationship into evidence about water ingress, moisture, ventilation, flues and electrical defects. The official installation audit made these mechanisms concrete for its historical installation cohort. The orchestrator also investigated how a repair would be performed and financed. The subsequent comparison considered the same identified defective project under funded contractual rectification, owner-funded replacement, paper completion and a pause without effective repair. Funded rectification ranked first. The Bill consequently links completion to repair actually performed and independently checked, with the payer, applicable coverage and necessary occupation arrangements identified.

From separate provisions to a combined world. The sixth comparison brought the dependencies together: a funded useful stage, valid covered repair, funded occupation until safe return, enforceable tenant protections and verified warmth and net financial benefit. The integrated statutory configuration and identical delivery under existing powers tied for first place; both ranked above the tested owner-gap and incomplete-delivery configurations. The outcome identifies the preferred delivery conditions independently of the legal route’s label. The proposed Bill provides their statutory expression while the matched comparison preserves equivalent treatment of the same achieved outcomes.

Question developed	Empirical distinction examined	Computational or legislative result
General policy course	Higher performance together with owner expenditure and tenant effects	Financing and execution become concrete research targets.
Funded delivery	Binding eligible finance, actual work, net tenant benefit and covered costs	The complete funded alternative is preferred within the revised set.
Legal implementation	Different legal routes with matched achieved consequences	Exact equality directs inquiry toward observable delivery differences.
Funding shortfall	A technically independent, safe and fully funded stage	Staging is selected on those conditions and expressed in clause 19.
Defect response	Available coverage and effective repair versus paperwork or delay	Funded rectification is selected and linked to verified completion.
Combined design	Finance, staging, repair, occupation and net tenant benefit together	Complete delivery ranks jointly first through both matched legal routes.

Table 1: The British case as a sequence of autonomous inquiry and normative computation. Each result is interpreted within its recorded comparison; the columns show what developed between tasks.

The published final proposal contains 34 clauses and four Schedules [12]. Its provisions connect project admission, consents, costs, contributions, procurement, occupation, staged work, verification, remedies and records. The autonomous cycle developed the substance of these connections. An eventual reimbursement required an arrangement for earlier payments; temporary accommodation required funded continuation through an overrun; a financial protection required an enforceable obligor and a funding source; and a repair promise required valid coverage and verified performance.

This case demonstrates sustained improvement through increasingly specific empirical analysis. The research moves from the performance of a housing stock to the causal and financial conditions of an executable project, then examines how those conditions interact. The shared ontology keeps the affected actors and consequences connected across the successive questions. Recalculation tests the resulting worlds under the same normative foundation, and the legislation expresses the conditions supported by that comparison.

5.4 Financial services: from an unknown condition to a research target

The financial-services study began with a business providing fiat and crypto-related services across jurisdictions. Its ontology-based route kernel represented the distribution of services, contractual roles and asset control. The task-specific decision operator combined legal conditions and economic comparisons. It reused JudgeAI ontology access, typed actor–action relationships, empirical identifiability checks and input receipts. Its corporate rules and financial comparison were separately specified; the TI/minimax operator for human actors remains the normative operator in the legislative studies; this financial application uses its specified corporate legal and economic comparison.

Here the computational trace has a particularly direct research function. A legal predicate whose applicability depends on an unknown company fact returns that unresolved condition together with the entity, rule and required fields. The kernel converts it into a research-queue item linked to the

rule's sources. Missing economic parameters generate corresponding requests. The output therefore identifies a task such as establishing a particular entity's ownership condition or obtaining a relevant cost parameter, rather than merely stating that further research is needed.

This is an implemented handoff from evaluation to inquiry. Whether the subsequent work is a company-fact check, a source investigation or a numerical analysis depends on the unresolved condition. Once a verified fact or a declared scenario assumption is supplied, the same condition can be evaluated again and its downstream consequences inspected. Unknown, assumed and established inputs remain distinguishable throughout the process.

5.5 A custody decision and the empirical boundary that determines it

The completed financial comparison evaluated 48 explicitly differentiated configurations. They varied the European base, the mode of European execution, and whether custody and Dubai brokerage functions were performed within the group or by external providers. A common commercial calibration was held fixed across the compared profiles. It combined researched legal and cost information with explicitly declared operating forecasts and price assumptions; it did not present these assumptions as actual partner quotations.

The custody comparison provides a concrete example of deeper analysis changing the content of the recommendation. At an annual client flow of USD 600 million and European assets under custody of USD 1 million, the selected configuration used an Irish principal operator and an own Dubai broker, with custody provided externally. In a paired profile with the same client flow and European assets under custody of USD 250 million, the selected configuration also brought European custody into the group.

The selection criterion in this study was annual after-tax economic result, including a specified charge for regulatory capital and principal liquidity. The change in custody affected provider charges, own operating expenses, capital and entity-level tax calculations. The result was obtained by evaluating the resulting configurations under the same accounting and selection rules. The question was not whether custody was attractive in isolation, but which allocation of services produced the preferred complete configuration for the specified profile.

The saved sensitivity analysis located the European custody transition at approximately USD 155.6 million of assets under custody, with annual client flow fixed at USD 600 million. The evaluated scenarios on the two sides of that transition differ in whether European custody is external or internal. A separate analysis under the same base profile placed the Dubai custody transition at approximately USD 202.6 million. These are boundaries within the declared scenarios and calibration.

The practical research question becomes more precise as a result. A business near the European boundary needs a better forecast of relevant custody balances and closer examination of the costs that move that boundary. The volume of payment flows by itself is insufficient. For a business far from the examined boundary, additional precision about that same quantity can have a different significance for the current selection. The calculation therefore supplies information that improves how the next empirical investigation can be directed.

5.6 When a changed input leaves the outcome unchanged

A further pair of financial profiles examined company ownership and Lithuanian tax relief. Direct individual ownership and corporate ownership led to different eligibility outcomes for the modelled zero-rate and reduced-rate cases. Yet the regulated company had a loss under both profiles. Its current corporate income tax was therefore zero in both. The profitable platform company still incurred USD 5,268 of tax, and the group economic result remained equal.

The trace distinguishes a changed applicability condition from a changed financial consequence. This matters for the quality of the recommendation: the system should not attribute a current tax saving

to an eligibility difference that produces none in the represented accounts. It also determines a more informative next inquiry. Ownership remains relevant, but the forecast of taxable income determines when that eligibility distinction becomes financially active.

Together, the examples show several outcomes of empirical development. Inquiry can strengthen the grounds of an existing normative choice; a refined operating profile can change the selected configuration; and a changed input can be shown to have no current effect on the final comparison. Traceability makes these outcomes distinguishable and gives each a concrete role in continued reasoning.

Empirical distinction	Computational finding	Consequence for inquiry or decision
Amsterdam’s governing policy course	The native comparison selects specific-impact regulation from five general approaches.	The package replaces a mandatory citywide overnight-stay range with conditions addressing local impacts and capacity.
Amsterdam financing incidence	Joint financial calibration supports a substantive preference for property-tax financing.	Investigate incidence and dependence across the affected actors; the financing basis is retained with a more developed empirical ground.
British project delivery	Six connected comparisons develop finance, staging, repair and the combined package.	The Bill links the resources and protections required for the represented benefits to occur.
European custody balances	At USD 600m annual flow, the selected configuration changes at approximately USD 155.6m of European custody assets.	Refine the custody forecast and the associated cost assumptions; distinguish external and own custody in the recommendation.
Ownership and tax eligibility	The eligibility branch changes, while current group tax and the economic result remain equal in the paired profiles.	Follow the condition through taxable income before assigning an economic advantage to the tax relief.

Table 2: Examples of the interaction in the completed studies. Each finding refers to its stated empirical model and decision operator. The financial quantities are conditional scenario results.

6 Relation to alignment, contextual judgment and adaptive inquiry

The comparison with contextual judgment concerns the mechanism through which empirical investigation can improve a normative conclusion. Anthropic’s July 2026 study, *Claude’s values across models and languages*, examines expressed value tendencies across models and languages. It places this work within the aim of cultivating context-sensitive good judgment, and reports structured variation after controlling for task, topic and user-expressed values [8]. Its object of study is expressed model behaviour. JudgeAI examines the explicit empirical-to-normative computation within an individual regulatory problem.

Anthropic’s constitution discusses a trade-off between clear decision procedures and contextual judgment: procedures support predictability and evaluation, while judgment can respond to circumstances that rules did not anticipate [9]. JudgeAI addresses the interaction between these requirements through a different arrangement. The empirical representation can develop as the orchestrator investigates the situation, while the normative comparison remains an explicit computational operation. Flexibility resides in the investigation and representation of reality as well as in the alternatives considered; reproducibility applies to each specified evaluation.

The architectural distinction concerns the relation between acquired knowledge and normative preference. In JudgeAI, that relation is available as an executed computation that the autonomous orchestrator can investigate and develop. In the interaction examined here, the engine exposes that relation and the orchestrator acts on it. A change of recommendation can be traced to a revised consequence estimate, a newly applicable condition or an additional alternative. This makes empirical responsiveness an operational property of the reasoning process.

Recent work also provides relevant concepts outside normative decision-making. Iris organises inquiry around an evolving information state, revisable claims and decision-critical unknowns, retaining candidate solutions using a fixed development evaluator [10]. Decision-focused active learning proposes selecting measurements for their expected contribution to a downstream industrial decision [11]. These works situate empirical inquiry as part of developing a solution. JudgeAI adds the substantive legal foundation that evaluates consequences for affected actors and compares regulatory worlds. The present article concerns the interaction made possible once such a foundation has an executable, traceable representation.

The July study and these neighbouring methods thus play different roles in the argument. The former motivates a comparison between contextual normative judgment and explicit normative computation. The latter help describe inquiry and information acquisition. JudgeAI's computational legal theory supplies the normative basis that the interaction requires. The contribution examined here is how that basis becomes a working instrument for continued empirical reasoning.

7 Scientific significance of the autonomous process

The cases establish several forms of development within a common architecture. Empirical inquiry can participate in selecting the governing policy course, strengthen the financial ground of a retained choice, introduce an executable alternative, distinguish an effective remedy from an administrative action and evaluate the conditions of a complete regulatory package. Each form is a change in the reasoning state that can be followed through the normative comparison.

This explains why traceability is productive. The trace identifies what the system should investigate, connects the new evidence to a particular represented consequence and allows the orchestrator to verify the effect of its revision. The ontology adds the semantic structure needed to make that inquiry precise. The computational legal theory gives its empirical findings normative significance. Their interaction enables autonomous progress from one evaluated state to the next.

For complex legislation, the shared representation also carries information across provisions. Financing, business admission, implementation and remedies may affect the same person through different roles and relationships. A burden removed by one measure can reappear through another; a promised benefit may depend on a resource or consent supplied elsewhere. The orchestrator can investigate these links and submit the combined world to the engine. The UK study demonstrates that progression from separate delivery questions to a calculation of their joint conditions. The Amsterdam study demonstrates its reach from governing policy through the incidence of package finance.

Improvement therefore has observable content in the research record: a more informative population or dependence structure; a better specified mechanism; an additional feasible alternative; verified transmission of evidence into the computation; or a more coherent set of conditions in the proposed law. An unchanged choice can be the result of a deeper examination, just as a revised choice can follow from a newly represented consequence. The trace makes the substantive reason visible in either case.

Deterministic evaluation supports this continuing development. A specified empirical state and engine version yield a reproducible calculation. Further inquiry changes the state submitted for evaluation while preserving the normative comparison. The process can continue through successive research

questions for as long as they develop consequential knowledge. Evidence from later implementation enters the same structure, allowing observed effects to refine the account on which subsequent decisions are based.

The resulting feedback is available during the formation of a regulation. It connects an autonomous agent's research capabilities to an explicit account of why a finding matters for normative choice. That is the central scientific value of making the theory of decision computable and fully traceable: the system gains a means of developing its own solution through increasingly detailed investigation of reality.

8 Conclusion

JudgeAI's autonomous orchestrator and computable theory of normative choice form a continuing empirical reasoning process. The engine makes the grounds of a decision explicit; the ontology identifies their empirical meaning; the orchestrator investigates them; and recalculation evaluates the developed worlds. Each stage provides material for the next.

The Amsterdam case shows this process reaching both the initial direction of tourism regulation and the joint financial incidence of its supporting measures. The England-and-Wales case shows successive questions becoming an integrated legislative design through six normative comparisons. The financial-services study further illustrates how traceable dependencies and decision boundaries give subsequent empirical inquiry a specific destination.

Traceability thus operates within autonomous decision formation. It enables the system to deepen the evidence, refine the consequences, develop the alternatives and improve the regulatory solution under a stable normative foundation. The more thoroughly the relevant reality is investigated and represented, the more substantively grounded the next normative evaluation can become.

Research materials

The article examines completed JudgeAI case records through 18 September 2026. The autonomous AI orchestrator performed the source investigation, empirical analysis, successive task formulation and interaction with the computational pipeline. The records retain the distinction between observations, source-model forecasts, declared comparison conditions and calculated results. The case calculations and conditional scenario comparisons are reported as results of the present article. Citations identify published theoretical work, public legislative texts and the external empirical sources used in the analysis; references to legislative texts accompany the corresponding provisions.

References

- [1] Kozlov, Y., Bajwa, T., and Shutova, M. *Law as Computation: An End-to-End System for Normative Decision-Making*. JudgeAI Research, 10 September 2026. <https://judgeai.space/moral-engine-ui/law-as-computation-2026.pdf>.
- [2] JudgeAI. *Consolidated Package of Draft Regulatory Acts: Tourism Activity in Amsterdam*. 11 September 2026. <https://judgeai.space/legislation/amsterdam-tourism-2026>.
- [3] Municipality of Amsterdam, Onderzoek en Statistiek. *Toeristische draagkracht van wijken in Amsterdam in 2025*. May 2026. <https://onderzoek.amsterdam.nl/publicatie/toeristische-draagkracht-van-wijken-in-amsterdam-in-2025>.
- [4] Municipality of Amsterdam, Onderzoek en Statistiek. *Bezoekersprognose Amsterdam 2026–2028*. May 2026. <https://onderzoek.amsterdam.nl/publicatie/bezoekersprognose-amsterdam-2026-2028>.

- [5] Statistics Netherlands (CBS). *Zelfstandigen; inkomen, vermogen, kenmerken*. StatLine table 84466NED. <https://www.cbs.nl/nl-nl/cijfers/detail/84466NED>.
- [6] Statistics Netherlands (CBS). *Vermogen van huishoudens; huishoudenskenmerken, vermogensbestanddelen*. StatLine table 83834NED. <https://www.cbs.nl/nl-nl/cijfers/detail/83834NED>.
- [7] APE. *Rapportage Amsterdam toeristenbelasting*. 2023, Tables 1 and 7. https://openresearch.amsterdam/image/2023/6/19/effect_toeristenbelasting_overnachtingen.pdf.
- [8] Kearney, M., et al. *Claude’s Values Across Models and Languages*. Anthropic, 13 July 2026. <https://www.anthropic.com/research/claude-values-models-languages>.
- [9] Anthropic. *Claude’s Constitution*. Version consulted 18 September 2026. <https://www.anthropic.com/constitution>.
- [10] Fu, S., et al. *Beyond Solution-Centric Search: Adaptive Inquiry and Knowledge Revision for Autonomous ML Engineering*. arXiv:2608.02143, 3 August 2026. <https://arxiv.org/abs/2608.02143>.
- [11] Srinivas, N., Ray, D., and Nakouzi, E. *Decision-Focused Active Learning for Scale-Aware Critical-Materials Recovery*. arXiv:2609.09413, 8 September 2026. <https://arxiv.org/abs/2609.09413>.
- [12] JudgeAI. *Private Rented Homes: Funded Energy Improvements. Draft Bill and Analysis*. 17 September 2026. <https://judgeai.space/legislation/uk-funded-energy-improvements-2026>.
- [13] Department for Energy Security and Net Zero (DESNZ). *Improving the energy performance of privately rented homes: impact assessment*. January 2026. GOV.UK, final impact assessment.
- [14] DESNZ. *Domestic Private Rented Sector Minimum Energy Efficiency Standards: Final Evaluation Report*. January 2026. Official evaluation report.
- [15] DESNZ. *National Energy Efficiency Data-Framework: Impact of Measures Data Tables 2026*. June 2026. GOV.UK, NEED 2026 tables.
- [16] Welsh Government. *The Impact on Health of the Welsh Government Warm Homes Schemes: Technical Report*. 2019. Official technical report.
- [17] DESNZ and Ofgem. *Solid Wall Insulation Installed under ECO4 and GBIS: Statistical Audit Results*. 2025, corrected November 2025. Official statistical audit.
- [18] DESNZ. *Warm Homes: Local Grant - Guidance for Local Authorities*. Updated 2 July 2026. <https://www.gov.uk/government/publications/warm-homes-local-grant>.