Task-Optimal Pidgin:
Information-Theoretic Specification Across Human–AI Expressiveness Gaps

Constantine Goltsev[1]

[1]Independent Researcher

July 2026 — Draft preprint

Abstract

As machine reasoning systems become more expressive than their human operators, natural-language requests can appear well posed to the sender while remaining compatible with a large family of machine-side interpretations. We model human–AI task specification as semantic communication between unequal representational apparatuses. A human apparatus and a machine apparatus induce, respectively, coarse and fine partitions of a latent intent space. The prompt is a costly codeword; the machine decodes it into an action; and semantic distortion is task regret rather than symbol error. We define residual intent entropy, ontology-refinement distance, task-relevant cognitive distance, and a human-cost-constrained expressibility frontier. Within this model we establish: (i) an information-theoretic lower bound on intent recovery; (ii) an exact ambiguity identity for nested ontologies; (iii) a minimal task-sufficient abstraction theorem that formalizes the physicist's "spherical cow"; (iv) a task-semantic rate–distortion converse; and (v) an optimal value-of-information rule for clarification questions. These results motivate SPEAR, a compact schema that combines natural language, mathematical notation, types, explicit abstraction choices, constraints, evaluation criteria, and clarification policies. Worked examples from systems engineering, statistics, statistical mechanics, and algebraic geometry illustrate why the optimal human–machine pidgin should be *maximally sufficient*, not maximally detailed. The paper is a theoretical position paper: the scalar "levels" of expressiveness are modeled as nested representational partitions, and empirical validation remains future work.

Central thesis. More capable decoding cannot reconstruct task-relevant information that was never encoded. A useful human–AI pidgin should therefore transmit the smallest task-sufficient abstraction at the lowest human cognitive cost, while triggering clarification whenever the expected value of another bit exceeds its interaction cost.

1 Introduction

Natural language is remarkably efficient among agents who share bodies, environments, conventions, and broadly similar cognitive machinery. Human conversation exploits common ground, pragmatic inference, and cooperative repair rather than spelling out every assumption (Grice, 1975; Clark and Brennan, 1991; Frank and Goodman, 2012). That economy becomes fragile when the receiver's hypothesis space is much larger than the sender realizes.

Consider a human with an illustrative representational "level 7" communicating with an AI system at "level 33." These numbers should not be interpreted as IQ, a universal intelligence scale, or a claim about current systems. They denote nested capacities to distinguish concepts, relations, abstractions, and possible task completions. A request such as "make the model faster" may denote one action in the human's active ontology, yet the machine can distinguish latency, throughput, compilation time, training time, energy, dollar cost, time-to-market, tail latency, and many other objectives. The machine's superior expressiveness does not itself select the intended one.

Classical information theory deliberately abstracts away meaning to study reliable symbol transmission (Shannon, 1948). Rate–distortion theory then asks how many bits are needed when approximate reconstruction is acceptable (Shannon, 1959). The information bottleneck asks for short representations that preserve information relevant to another variable (Tishby et al., 1999). Semantic communication, signaling games, and probabilistic pragmatics address related questions about meaning and coordinated action (Lewis, 1969; Juba and Sudan, 2008; Bao et al., 2011; Frank and Goodman, 2012). Meanwhile, machine-learning practice has exposed a broad failure mode in which many solutions fit an observed specification but behave differently outside it (D'Amour et al., 2022).

This paper combines these ideas around a narrow question:

> What intermediate language minimizes task error when a cognitively bounded human communicates with a much more expressive machine?

The answer proposed here is not "write everything down." More detail raises human cost, can obscure the decision-relevant structure, and can create false precision. The target is instead a *task-optimal semantic compression*: preserve the invariants that determine acceptable action and discard the rest. This is the formal content of the "spherical cow" move in physics.

Contributions.

1. A partition-based model of asymmetric representational apparatuses and a distinction between raw ontology distance and task-relevant cognitive distance.
2. Five theorem-level statements connecting underspecification, conditional entropy, task-sufficient abstraction, rate–distortion, and clarification.
3. A human-cost-constrained expressibility frontier that characterizes the optimal pidgin as a Pareto object rather than a maximally formal language.
4. SPEAR, a proposed schema for mixed natural/symbolic specification.
5. Worked examples and falsifiable experimental predictions.

Status of the results. The information lower bounds are adaptations of classical results. The ontology and abstraction propositions are elementary consequences of the definitions introduced here. They should not be mistaken for empirical evidence that human or machine expressiveness is one-dimensional.

2 Communication Model

2.1 Latent intent, messages, and actions

Let Θ be a random variable on a finite or countable intent space Ω. An intent contains everything that could matter to the user's desired result: objective, object types, assumptions, constraints, tolerances, acceptable outputs, and context. All logarithms use a common base, and entropy units follow that choice. The human emits a message M using a language $\mathcal{L}$:

$$M \sim e_{\mathcal{L}}(\cdot \mid \Theta, C),$$

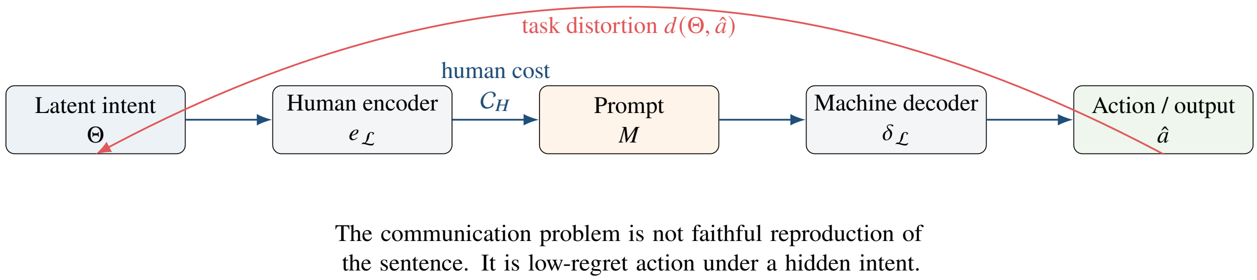

Figure 1: Human–AI task specification as a semantic communication channel.

where C denotes shared conversational and environmental context. The machine chooses an action or output

$$\hat{a} = \delta_{\mathcal{L}}(M, C).$$

A task distortion function

$$d : \Omega \times \mathcal{A} \to \mathbb{R}_{\geq 0}$$

measures the loss from taking action a under intent θ. Distortion may be exact semantic mismatch, expected regret, safety loss, scientific error, or violation of acceptance tests.

The human pays a production and comprehension cost $C_H(M)$. It may combine message length, formal-symbol burden, working-memory depth, time, and probability of authoring error. The correct cost is empirical and user-dependent; no universal token-to-cognition conversion is assumed.

2.2 Representational apparatuses as partitions

A representational apparatus induces distinctions on Ω. Let Π_H be the partition of intent space available to the human in the current interaction and Π_A the machine's finer partition. Write

$$Z_H = \pi_H(\Theta), \qquad Z_A = \pi_A(\Theta).$$

If every AI cell lies inside a human cell, then Π_A refines Π_H, written $\Pi_A \preceq \Pi_H$. A "level k" apparatus can be modeled by a nested sequence

$$\Pi^{(1)} \succeq \Pi^{(2)} \succeq \cdots \succeq \Pi^{(k)},$$

with higher levels supporting finer distinctions. Equivalently, the levels generate nested sigma-algebras.

The raw ontology-refinement distance is

$$D_{\text{ont}}(H, A) = \mathrm{H}(Z_A \mid Z_H). \tag{1}$$

It measures how many machine-side distinctions remain unresolved after the human-level description is known.

Raw distance is not yet the communication problem. Many machine distinctions may be irrelevant to the requested action. Let Z_T denote a task-relevant latent variable, such as the correct action class, acceptable output class, or predictive quantity. Define

$$D_{\text{task}}(H, A) = \mathrm{H}(Z_T \mid Z_H). \tag{2}$$

A large D_{ont} can coexist with a small D_{task}: the AI may know thousands of irrelevant distinctions. The pidgin need only bridge the task-relevant gap.

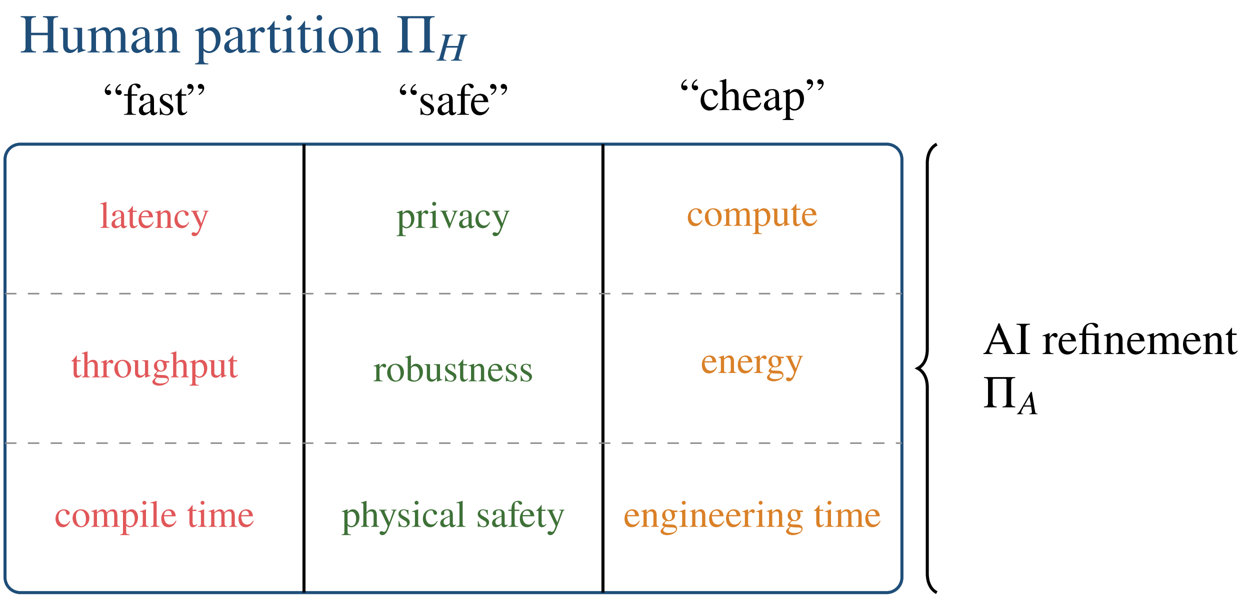

Figure 2: A coarse human category may contain many machine-side interpretations. Only distinctions that change the desired action contribute to task-relevant cognitive distance.

2.3 Residual intent entropy and interpretive expansion

For a realized prompt m, define residual intent entropy

$$\mathcal{A}(m) = \mathrm{H}(\Theta \mid M = m, C),$$

and task residual entropy

$$\mathcal{A}_T(m) = \mathrm{H}(Z_T \mid M = m, C).$$

A system can produce fluent output while $\mathcal{A}_T(m)$ remains high. Coherence is conditional on a selected interpretation; it is not evidence of intent fidelity.

For finite support, define the interpretive expansion factor

$$\Gamma(m) = \frac{|\operatorname{supp}(Z_A \mid m)|}{|\operatorname{supp}(Z_H \mid m)|}.$$

This cardinality diagnostic is crude but intuitive: a prompt can feel unique in the sender's active ontology while expanding into many candidates in the receiver's ontology.

3 Information-Theoretic Results

3.1 Missing task information cannot be decoded

Theorem 1 (Intent recovery lower bound)**.** *Let Z_T take values in a finite set of size $N \geq 2$, and let $\widehat{Z}_T$ be any estimate based on (M, C). Its probability of error $P_e = \Pr[\widehat{Z}_T \neq Z_T]$ satisfies*

$$P_e \geq \max\left\{0, \frac{\mathrm{H}(Z_T) - \mathrm{I}(Z_T; M, C) - \log 2}{\log N}\right\}. \tag{3}$$

If Z_T is uniform, then $\mathrm{H}(Z_T) = \log N$ *and this becomes*

$$P_e \geq \max\left\{0, 1 - \frac{\mathrm{I}(Z_T; M, C) + \log 2}{\log N}\right\}.$$

Consequently, increasing decoder capability cannot drive P_e to zero when the message and shared context leave substantial conditional entropy about the task class.

Proof. Fano's inequality gives $\mathrm{H}(Z_T \mid M, C) \leq \log 2 + P_e \log N$. Substituting $\mathrm{H}(Z_T \mid M, C) = \mathrm{H}(Z_T) - \mathrm{I}(Z_T; M, C)$ and rearranging proves equation (3) (Fano, 1961). The decoder's computational sophistication does not alter the information available in (M, C). □

Remark 1. A strong prior can improve average accuracy, but it does so by guessing likely intents. It does not remove the ambiguity of an individual underspecified request. The distinction is

$$\text{high posterior probability} \neq \text{semantic identification.}$$

3.2 Ambiguity induced by ontology refinement

Theorem 2 (Refinement ambiguity identity)**.** *Assume $\Pi_A \preceq \Pi_H$, the message reveals exactly Z_H, and conditional on $Z_H = h$, the $r(h)$ compatible AI cells are equiprobable. Then*

$$\mathrm{H}(Z_A \mid M) = \mathbb{E}_{Z_H}[\log r(Z_H)]. \tag{4}$$

If an additional constraint field Q is supplied, the remaining ambiguity is

$$\mathrm{H}(Z_A \mid M, Q) = \mathrm{H}(Z_A \mid M) - \mathrm{I}(Z_A; Q \mid M). \tag{5}$$

Proof. Because M determines Z_H, conditional entropy decomposes over human cells. Uniformity gives $\mathrm{H}(Z_A \mid Z_H = h) = \log r(h)$; averaging proves equation (4). Equation equation (5) is the defining identity for conditional mutual information. □

The theorem assigns a local value to a prompt field: its semantic benefit is the conditional mutual information it contributes. Under additive cognitive cost, a natural one-step design score is

$$\rho(Q \mid m) = \frac{\mathrm{I}(Z_T; Q \mid M = m, C)}{\mathbb{E}[C_H(Q) \mid m]}. \tag{6}$$

Fields with high ρ should be elicited first, subject to dependencies among fields.

3.3 The spherical-cow theorem

A good abstraction discards details that do not change the task-relevant distribution. Let Y be the desired task variable. Define an equivalence relation on intents:

$$\theta \sim_T \theta' \quad \Longleftrightarrow \quad p(Y \mid \Theta = \theta) = p(Y \mid \Theta = \theta'). \tag{7}$$

Let $Z_T^* = \Theta/{\sim_T}$ be the quotient label.

Theorem 3 (Minimal task-sufficient abstraction)**.** *The quotient Z_T^* is a deterministic sufficient statistic for Y:*

$$Y \perp \Theta \mid Z_T^*, \qquad \textit{equivalently} \qquad \mathrm{I}(Y; \Theta \mid Z_T^*) = 0. \tag{8}$$

Moreover, every deterministic statistic $Z = \phi(\Theta)$ satisfying $Y \perp \Theta \mid Z$ must refine Z_T^*. Thus Z_T^* has minimum cardinality among deterministic task-sufficient abstractions.

Proof. By construction, all intents in one quotient class induce the same conditional law of Y, proving conditional independence. Suppose a sufficient statistic Z maps two intents with different conditional laws of Y to the same value. Conditioning on that value would leave the law of Y dependent on which intent occurred, contradicting sufficiency. Hence every sufficient Z separates distinct quotient classes and therefore refines Z_T^*. □

Corollary 4 (Formal spherical cow)**.** *A detail of the world may be omitted with zero task-semantic loss exactly when it varies only within classes of $\sim_T$. The optimal zero-loss abstraction is not the most detailed model; it is the coarsest model that preserves the task-relevant conditional law.*

This theorem turns a physics heuristic into a design criterion. Replacing a cow by a sphere is justified only if limb geometry, local fur variation, and posture do not change the target quantity beyond the accepted distortion. Correct mathematics applied to an insufficient abstraction remains efficiently wrong.

3.4 Task-semantic rate–distortion

For general approximate tasks, define

$$R_T(D) = \inf_{\substack{p(m|\theta,c),\,\delta: \\ \mathbb{E}[d(\Theta,\delta(M,C))]\le D}} \mathrm{I}(\Theta; M \mid C). \tag{9}$$

This is a task-semantic rate–distortion function: the minimum intent information required to keep expected task loss below D.

Theorem 5 (Task-semantic converse)**.** *Any human–machine language and decoder achieving expected distortion at most D must satisfy*

$$\mathrm{I}(\Theta; M \mid C) \ge R_T(D). \tag{10}$$

If the effective information conveyed by the human prompt is below $R_T(D)$, no increase in machine reasoning power can achieve the target distortion from that prompt alone.

Proof. The first inequality follows directly from the infimum defining $R_T(D)$, in the same spirit as Shannon's fidelity-constrained source coding converse (Shannon, 1959). The second statement follows because every decoder based only on (M, C) lies in the feasible decoder class over which the bound is defined. □

3.5 When should the machine ask?

Let the posterior Bayes risk after message m be

$$\mathcal{R}(m) = \min_{a\in\mathcal{A}} \mathbb{E}[d(\Theta, a) \mid M = m, C]. \tag{11}$$

A clarification query Q has interaction cost $c(Q)$ and expected post-query risk

$$\mathcal{R}(m; Q) = \mathbb{E}_{q|m}\left[\min_{a} \mathbb{E}[d(\Theta, a) \mid m, q, C]\right].$$

Theorem 6 (Optimal clarification rule)**.** *For a single optional clarification step, asking Q is optimal exactly when*

$$\mathrm{VOI}(Q \mid m) := \mathcal{R}(m) - \mathcal{R}(m; Q) > c(Q). \tag{12}$$

Among available queries, the optimal one maximizes $\mathrm{VOI}(Q \mid m) - c(Q)$.

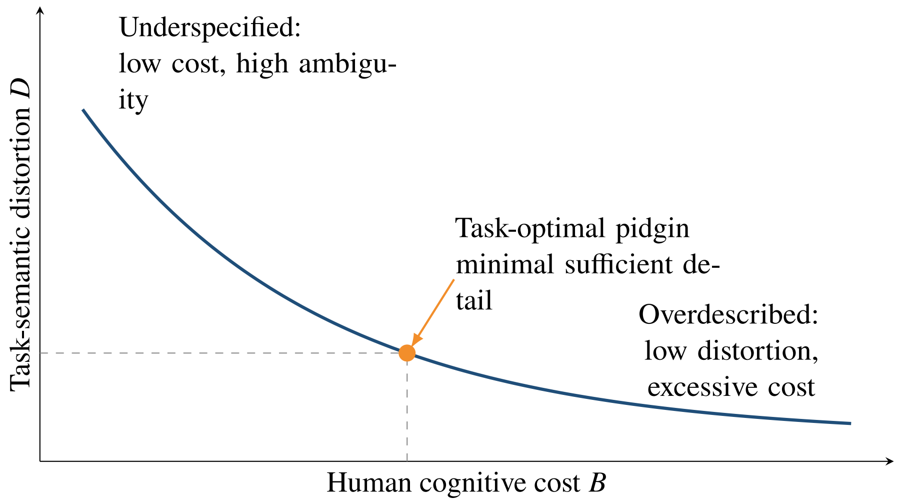

Figure 3: The pidgin is selected on a human-cost versus task-distortion frontier. The optimum depends on task stakes, user skill, context, and interaction cost.

Proof. The expected total loss of acting immediately is $\mathcal{R}(m)$. The expected total loss of asking and then acting is $c(Q) + \mathcal{R}(m; Q)$. Comparing the two yields equation (12), matching the classical value-of-information principle (Howard, 1966). □

This turns "ask when unclear" into a decision rule. A machine should not ask for every omitted detail. It should ask when the detail is expected to change the action enough to justify the interruption.

4 The Pidgin Efficiency Frontier

A language that transmits every detail is unusable; a language that is easy but ambiguous is unsafe. Define the human-budget expressibility function

$$\mathcal{E}(B) = \sup_{\mathcal{L}, e_{\mathcal{L}}:\ \mathbb{E}[C_H(M)] \le B} \mathrm{I}(Z_T; M \mid C). \tag{13}$$

It is the maximum task-relevant information a human can convey under cognitive budget B. An optimal pidgin lies on the Pareto frontier between human cost and semantic distortion.

A convenient Lagrangian form is

$$\mathcal{L}^* \in \arg\min_{\mathcal{L}, e, \delta} \left\{ \mathbb{E}[d(\Theta, \delta(M))] + \lambda \mathbb{E}[C_H(M)] + \mu \mathbb{E}[C_{\text{int}}] \right\}, \tag{14}$$

where C_{int} counts clarification and repair. This parallels rate–distortion, information bottleneck, and minimum-description-length principles (Tishby et al., 1999; Rissanen, 1978).

One can also define an efficiency ratio

$$\eta(\mathcal{L}) = \frac{\mathrm{I}(Z_T; M \mid C)}{\mathbb{E}[C_H(M)] + \varepsilon},$$

but maximizing η alone is insufficient: a nearly empty message may have a deceptively favorable ratio. It should be optimized subject to a distortion or risk threshold.

Field	Information-theoretic role	Typical failure prevented
TASK	Names the decision class and establishes common ground	Answering a neighboring but different problem
OBJECTS & TYPES	Narrows valid operations and dimensions	Scalar/vector, data/model, set/sequence confusion
ABSTRACTION	Declares preserved invariants and discarded degrees of freedom	False precision or omitted causal variables
OBJECTIVE	Defines the task-relevant quotient or loss	Optimizing latency when throughput was intended
CONSTRAINTS	Restricts the feasible action set	Invalid, unsafe, unaffordable solutions
UNCERTAINTY	Supplies priors, unknowns, and confidence requirements	Treating assumptions as facts
OUTPUT	Defines representation and granularity	Correct analysis in unusable form
EVALUATION	Makes distortion observable	No acceptance criterion; unverifiable completion
INTERACTION	Encodes clarification threshold and default behavior	Guessing when a cheap question would resolve high risk
EXAMPLES	Defines category boundaries by contrast	Misreading edge cases or tacit style constraints

Table 1: SPEAR fields and the ambiguities they are intended to reduce.

5 SPEAR: A Proposed Shared Specification Pidgin

We propose **SPEAR**: *Shared Pidgin for Expressive Abstraction and Requirements*. It is not a new programming language. It is a compact schema that can be written in ordinary prose, YAML-like blocks, mathematical notation, or a graphical interface. Each field exists because it reduces a distinct component of task entropy.

5.1 Core schema

```
SPEAR/0.1

TASK:
  <verb + target object + decision/use context>

OBJECTS:
  <name> : <type, domain, units, shape, ownership>

ABSTRACTION:
  preserve: <task-relevant invariants>
  ignore:   <degrees of freedom intentionally discarded>
  assume:   <model family and boundary conditions>

OBJECTIVE:
  minimize|maximize|estimate|prove: <mathematical or verbal target>

CONSTRAINTS:
  hard: <must hold>
  soft: <preference> @weight=<number or ordering>

UNCERTAINTY:
  known: <observations>
```

```
  unknown: <latent variables>
  prior: <distribution or qualitative prior>

OUTPUT:
  type: <proof | code | table | design | posterior | decision>
  format: <schema, length, notation, audience>

EVALUATION:
  metric: <task distortion / acceptance tests>
  threshold: <acceptable error or regret>

INTERACTION:
  ask_if: expected_regret_from_ambiguity > <threshold>
  otherwise: state assumptions and proceed

EXAMPLES:
  positive: <one intended instance>
  negative: <one near-miss>
```

5.2 Why mixed natural and mathematical language?

Natural language carries goals, context, analogy, and exceptions cheaply. Formal notation carries types, invariants, constraints, and compositional structure compactly. The pidgin should exploit both. For example,

$$x^* = \arg\min_{x \in \mathcal{X}} L(x) \quad \text{s.t.} \quad g_j(x) \leq 0$$

eliminates several syntactic ambiguities, while prose must still say what x, L, and the constraints mean in the world.

False-precision warning. Formal notation compresses a model; it does not certify the model. A perfectly typed optimization problem can optimize the wrong objective, omit the dominant variable, or encode a politically contested value as if it were a physical constant. SPEAR therefore requires an explicit ABSTRACTION and EVALUATION layer.

6 Worked Prompt Transformations

6.1 Systems engineering: "Make it faster"

Loose request.

> Make the anomaly detector faster without hurting quality.

The machine can interpret "faster" as p50 latency, p99 latency, throughput, deployment time, or cost per event. "Quality" may mean recall, precision, alert volume, ranking stability, or downstream incident time.

SPEAR version.

```
TASK:
  Redesign the streaming anomaly detector for production deployment.

OBJECTS:
```

```
  stream_rate : Real[events/s]
  latency_p99 : Real[ms]
  score(x)    : Real
  label(x)    : {normal, anomaly}

ABSTRACTION:
  preserve: anomaly ranking above operational threshold
  ignore: exact score calibration below threshold
  assume: event distribution stationary over a 24 h calibration window

OBJECTIVE:
  minimize latency_p99

CONSTRAINTS:
  hard: throughput >= 1e6 events/s
  hard: recall_at_fixed_alert_budget >= baseline - 0.01
  hard: memory <= 64 GB
  soft: minimize cloud_cost @weight=secondary

OUTPUT:
  type: architecture + benchmark plan + pseudocode

EVALUATION:
  metric: (latency_p99, recall_at_budget, cost_per_1M_events)
  threshold: ship only if all hard constraints pass

INTERACTION:
  ask_if: the baseline workload or alert budget is missing
```

The schema identifies the task quotient: detector variants are equivalent if they preserve the operational ranking and constraints, even if their sub-threshold scores differ.

6.2 The spherical cow: heat loss

Loose request.

> Estimate how much heat a cow loses outside in winter.

A maximally detailed model would include anatomy, fur, wind, moisture, metabolism, radiation, posture, and ground contact. That detail is unnecessary for a first-order convective estimate.

SPEAR version.

```
TASK:
  Estimate steady-state convective heat loss to first order.

OBJECTS:
  cow := Sphere(radius r = 0.80 m)
  T_body = 38 C
  T_air  = -5 C
  h      = 10 W/(m^2 K)

ABSTRACTION:
  preserve: total exposed area and mean temperature difference
  ignore: limbs, local curvature, fur anisotropy, radiation, evaporation
```

```
  assume: uniform surface temperature; constant h; no ground contact

OBJECTIVE:
  estimate Qdot = h * 4*pi*r^2 * (T_body - T_air)

OUTPUT:
  type: point estimate + sensitivity to r and h

EVALUATION:
  threshold: order-of-magnitude accuracy only
```

The approximation is valid at the declared abstraction level. The relevant question is not whether a cow is spherical, but whether the discarded geometry changes the accepted decision.

6.3 Statistics: "Did the treatment work?"

Loose request.

> Analyze the data and tell me whether the treatment worked.

The word "worked" can denote a causal average effect, a subgroup effect, a predictive association, clinical significance, or a statistically significant difference.

SPEAR version.

```
TASK:
  Estimate the average causal treatment effect at 30 days.

OBJECTS:
  Y(1), Y(0) : potential outcomes in [0, 100]
  T          : {0,1} treatment indicator
  X          : pre-treatment covariates

ABSTRACTION:
  preserve: mean counterfactual difference
  ignore: effects after day 30
  assume: SUTVA; positivity; Y(1),Y(0) independent of T conditional on X

OBJECTIVE:
  estimate tau = E[Y(1) - Y(0)]

UNCERTAINTY:
  report: 95% interval and sensitivity to unmeasured confounding

EVALUATION:
  clinical threshold: tau >= 5 points
  statistical threshold: posterior P(tau >= 5 | data) >= 0.95
```

Here the mathematics defines the estimand, while the abstraction block exposes the assumptions that make it identifiable.

6.4 Statistical mechanics: low-temperature heat capacity

Loose request.

> Derive the heat capacity of a crystal at low temperature.

SPEAR version.

```
TASK:
  Derive the leading low-temperature heat capacity of a 3D insulating solid.

OBJECTS:
  omega(k) = v_s * |k|            # acoustic dispersion
  g(omega) proportional to omega^2
  beta = 1/(k_B*T)

ABSTRACTION:
  preserve: long-wavelength acoustic modes and density-of-states scaling
  ignore: optical modes, anharmonicity, disorder, electronic excitations
  assume: isotropic continuum; harmonic phonons; T << Theta_D

OBJECTIVE:
  derive the leading power law and prefactor from
  U = integral_0^{omega_D} hbar*omega*g(omega)/(exp(beta*hbar*omega)-1) d omega

OUTPUT:
  type: derivation showing C_V proportional to T^3
  format: graduate statistical-mechanics notation
```

The prompt is compact because the correct abstraction preserves the infrared modes that control the asymptotic law.

6.5 Algebraic geometry: singular locus

Loose request.

> Find the bad points of this algebraic surface.

SPEAR version.

```
TASK:
  Compute the singular locus of an affine hypersurface.

OBJECTS:
  k : algebraically closed field, char(k)=0
  f in k[x_1,...,x_n]
  X = V(f) subset A_k^n

ABSTRACTION:
  preserve: scheme-theoretic vanishing conditions of the Jacobian ideal
  ignore: projective points at infinity

OBJECTIVE:
  compute Sing(X) = V(f, partial_1 f, ..., partial_n f)

OUTPUT:
  type: defining ideal + dimension + decomposition if reducible

INTERACTION:
```

```
ask_if: characteristic or affine/projective setting is unspecified
```

This example shows why typed mathematical notation can function as a high-bandwidth pidgin across domains.

7 Design Principles Derived from the Model

The formalism yields the following practical criteria.

P1. Transmit the task quotient, not the world. Describe distinctions that change acceptable action; omit those that do not.

P2. Name the distortion. A prompt is not well posed until "wrong" can be measured or at least operationally recognized.

P3. Type the objects. Types eliminate entire classes of invalid interpretations at low cognitive cost.

P4. Declare the abstraction boundary. State both what is preserved and what is intentionally ignored.

P5. Separate hard constraints from soft preferences. Otherwise the decoder must invent tradeoffs.

P6. Expose assumptions and priors. Hidden assumptions are compressed bits supplied by the machine's default world model.

P7. Use examples near decision boundaries. A positive example and a near-miss often yield more conditional information than another paragraph of generic prose.

P8. Clarify by value of information. Ask the cheapest question expected to change the action most.

P9. Match formalism to stakes. High-risk tasks justify larger human cost B; routine tasks may rationally rely on priors.

P10. Permit repair. Grounding is iterative; the pidgin is a protocol, not merely a sentence.

8 Failure Modes and Limits

8.1 The ontology need not be nested

Human and machine categories may cross-cut rather than refine one another. A domain expert can possess distinctions absent from a general model, while the model possesses broad distinctions outside the expert's focus. In that case D_{ont} should be replaced by a symmetric discrepancy between partitions or by conditional entropies in both directions.

8.2 The user may not know the true intent

Intent is often constructed during interaction. The latent-variable model should then be extended to a dynamic process Θ_t, with questions helping the user discover preferences rather than merely reveal fixed ones.

8.3 Wrong loss functions create specification gaming

A precise metric can induce behavior that satisfies the metric while violating the underlying purpose. SPEAR reduces syntactic ambiguity but cannot guarantee normative alignment. The EVALUATION field itself requires scrutiny.

8.4 More notation can reduce actual information

Symbols unfamiliar to the user increase authoring errors and may conceal uncertainty. Human cost must include probability of malformed specification, not just length. The same expression can be high-bandwidth for a physicist and noise for a novice.

8.5 Power asymmetry and contestability

A shared pidgin can improve precision while shifting power toward whoever defines the types, metrics, and abstraction boundary. High-stakes uses require provenance, appeal, and explicit uncertainty, not merely efficient communication.

9 Empirical Program

The theory suggests a benchmark in which participants communicate hidden tasks to models under controlled ontology gaps.

Task construction. Generate intent families with known task quotient Z_T^*, nuisance dimensions, and multiple plausible interpretations. Vary the ratio between task-relevant and irrelevant machine distinctions.

Conditions. Compare free-form natural language, natural language plus examples, SPEAR, and a fully formal specification. Measure human authoring time, error rate, model task regret, clarification count, and subjective burden.

Predictions.

1. Task error should track $\mathrm{H}(Z_T \mid M, C)$ more closely than raw prompt length.
2. Structured fields should produce the largest gains when they have high conditional information per human cost, as in equation (6).
3. Performance as a function of detail should be U-shaped once human authoring error and overconstraint are included.
4. Explicit ABSTRACTION fields should improve transfer to altered environments more than format-only structure.
5. Clarification governed by estimated value of information should dominate both "never ask" and "always ask" policies under mixed task stakes.

10 Discussion

Human language already behaves like lossy semantic compression. Speakers omit details because listeners infer them from common ground and shared priors (Grice, 1975; Clark and Brennan, 1991; Frank and Goodman, 2012). The human–AI case differs quantitatively and sometimes qualitatively: the machine may entertain many more completions, may lack embodied common ground, and may generate a polished answer before the ambiguity becomes visible.

The proposed pidgin is therefore not a replacement for natural language. It is a scaffold for moments where the task-relevant entropy is high. Its most important move is to force an abstraction choice: which variables, invariants, and distinctions are allowed to survive compression? In this sense, mathematical notation is valuable not because it is intrinsically superior, but because it offers a mature library of typed, compositional, low-ambiguity abstractions.

The optimal pidgin should evolve with the pair. Repeated interaction can establish conceptual pacts, defaults, and user-specific cost models. A novice may use prose plus examples; a domain expert may use equations; the same user may move between both. The language should therefore be adaptive and interface-mediated rather than fixed.

11 Conclusion

When a human at an illustrative "level 7" addresses a machine at "level 33," the goal is not to lift the human to level 33 or force the machine down to level 7. It is to construct the smallest shared representation that preserves the action-relevant structure of the task.

In information-theoretic terms, the design target is:

$$\boxed{\text{minimize human cost and task distortion} \quad \text{subject to task-sufficient mutual information.}}$$

The central engineering slogan is correspondingly simple:

Specify the invariants, not the universe.

A A Compact EBNF Sketch for SPEAR

```
specification ::= header section+
header        ::= "SPEAR/" version
section       ::= task | objects | abstraction | objective |
                  constraints | uncertainty | output |
                  evaluation | interaction | examples

task          ::= "TASK:" statement
objects       ::= "OBJECTS:" object_decl+
object_decl   ::= identifier ":" type [domain] [units] [shape]

abstraction   ::= "ABSTRACTION:" preserve ignore assumptions
preserve      ::= "preserve:" item+
ignore        ::= "ignore:" item+
assumptions   ::= "assume:" item+

objective     ::= "OBJECTIVE:" operator expression
operator      ::= "minimize" | "maximize" | "estimate" |
                  "prove" | "classify" | "design"

constraints   ::= "CONSTRAINTS:" (hard_constraint | soft_constraint)+
hard_constraint ::= "hard:" proposition
soft_constraint ::= "soft:" proposition ["@weight=" weight]

interaction   ::= "INTERACTION:" ask_rule default_rule
ask_rule      ::= "ask_if:" risk_expression comparison threshold
default_rule  ::= "otherwise:" statement
```

B Notation Summary

Symbol	Meaning
Θ	latent human intent
M	prompt or message
C	shared context
$\hat{a}$	machine action or output
$d(\theta, a)$	task-semantic distortion or regret
$C_H(m)$	human cognitive cost of message m
Z_H, Z_A	human and machine ontology cells
Z_T	task-relevant latent class
D_{ont}	raw ontology-refinement distance
D_{task}	task-relevant cognitive distance
$R_T(D)$	task-semantic rate–distortion function
$\mathcal{E}(B)$	task expressibility under human budget B

Acknowledgment and authorship note

This draft was developed through human–AI dialogue. The named human author is responsible for reviewing, validating, revising, and deciding whether any version is suitable for dissemination. The AI system is not listed as an author.

References

Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3–4):379–423, 623–656, 1948.

Claude E. Shannon. Coding theorems for a discrete source with a fidelity criterion. In *IRE National Convention Record*, volume 4, pages 142–163, 1959.

Robert M. Fano. *Transmission of Information: A Statistical Theory of Communications*. MIT Press, Cambridge, MA, 1961.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, pages 368–377, 1999. arXiv:physics/0004057.

Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

H. Paul Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics, Volume 3: Speech Acts*, pages 41–58. Academic Press, New York, 1975.

Herbert H. Clark and Susan E. Brennan. Grounding in communication. In Lauren B. Resnick, John M. Levine, and Stephanie D. Teasley, editors, *Perspectives on Socially Shared Cognition*, pages 127–149. American Psychological Association, Washington, DC, 1991.

Michael C. Frank and Noah D. Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998, 2012.

Alexander D'Amour, Katherine Heller, Dan Moldovan, and others. Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research*, 23(226):1–61, 2022.

Ronald A. Howard. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26, 1966.

Jie Bao, Prithwish Basu, Mike Dean, Craig Partridge, Ananthram Swami, William Leland, and James A. Hendler. Towards a theory of semantic communication. In *2011 IEEE Network Science Workshop*, pages 110–117, 2011.

Brendan Juba and Madhu Sudan. Universal semantic communication I. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 123–132, 2008.

David K. Lewis. *Convention: A Philosophical Study*. Harvard University Press, Cambridge, MA, 1969.