

How Far Can the Compression *See*?

Intelligence is compression that predicts. The test of a model is not how much it discards, but how long what remains keeps being right.

John Rector · August 16, 2026 · Figures current to July 2026 data · johnrector.me

ABSTRACT

A long tradition in mathematics and information theory connects intelligence with compression. A system that understands a body of information does not memorize every detail; it finds the pattern that generated those details and keeps the pattern instead. But compression alone is not intelligence. Three hundred million Americans can be compressed into the single word *people* — enormous compression, almost no intelligence. The intelligent representation is not the shortest description. It is the shortest description that continues to predict actual outcomes.

This essay uses the monthly U.S. household survey as a working example of an intelligent compression under strain. The Current Population Survey reduces a nation to about 74,000 sampled addresses, roughly 60,000 eligible households, a July 2026 response rate of 65.3 percent, and finally to three labor force categories: employed, unemployed, not in the labor force. That compression has been extraordinarily useful. It is now encountering a structure it was not built to preserve — a person who is simultaneously an employee and, with AI doing the work an organization used to do, the operator of several businesses.

The essay proposes prediction horizon as the practical measure of intelligence: how far a compressed representation can travel before reality breaks it. By that measure, a model is not judged by its authority, its sample size, or the elegance of its method. It is judged by the distance between what its categories predict and what subsequently becomes *actual*.

01 Intelligence Is Compression That Predicts

A strong tradition in mathematics, information theory, and computer science connects intelligence with compression. The underlying idea is simple. A system that understands a body of information does not need to remember every individual detail. It discovers the pattern that generated those details and represents the pattern instead.

One thousand points become **a line**.

Millions of planetary observations become **a small collection of equations**.

An entire language becomes **a grammar**.

Compression removes information. Intelligence determines which information can be removed without losing the ability to predict what happens next.

That final qualification matters. Compression alone is not intelligence. I can compress the entire population of the United States into one category:

People.

Three hundred million people have now become one word. The compression is extraordinary. The intelligence is almost zero. *People* preserves too little structure to predict anything useful about what those people will do.

The intelligent representation is not simply the shortest description. It is **the shortest description that continues to predict actual outcomes.**

02 Compressing America

Consider the monthly household survey that produces the unemployment rate.

The United States contains more than 300 million people living inside radically different economic realities. Some are billionaires. Some live below the poverty line. Some receive salaries. Some work by the hour. Some own several companies. Some are medically disabled. Some are retired and traveling. Some are building businesses while remaining employed. Some survive through several irregular income streams that do not resemble a conventional job.

The Current Population Survey does not observe all of them.

Each month, the Census Bureau starts with a probability sample of roughly 74,000 assigned addresses, of which about 60,000 are occupied and eligible for interview. In July 2026, the survey's response rate was **65.3 percent** — not a record low, but well beneath the 80 percent threshold at which federal statistical standards require a formal nonresponse bias analysis, and far from the roughly 87 percent the survey achieved a generation ago. That implies something in the neighborhood of 39,000 completed household interviews. The Bureau of Labor Statistics does not publish a monthly count of interviews, so that figure is derived rather than reported; the survey's own methodology handbook still describes about 54,000 completed interviews as normal, a number the handbook has not caught up on.

Census then weights those responses to represent the larger American population. The process ends by placing every person aged 16 and over into one of three categories: **employed**, **unemployed**, or **not in the labor force**.

FIGURE 01

The compression cascade: a nation reduced to three categories

U.S. civilian noninstitutional population, 16 and over	275,282,000
Addresses assigned to the monthly sample	~74,000
Households occupied and eligible for interview	~60,000
Interviews completed at a 65.3% response rate (derived)	~39,000
Weighted statistical model of the country	1 model
Published labor force categories	3

Each step discards information in exchange for tractability. The final step is the largest reduction and the one this essay is about. Population and response-rate figures: BLS Employment Situation, July 2026 (released August 7, 2026) and CPS response rate series LNU09300000. Sample-size figures: BLS Handbook of Methods, CPS Design, and Census Bureau CPS methodology. The ~39,000 figure is derived from the published response rate, not reported by BLS.

Three hundred million people become 74,000 addresses. Those become roughly 60,000 eligible households. Those become about 39,000 responding households. Those become a weighted model of the country. That model becomes three labor force categories.

This is compression. It is also intelligence.

The survey discards nearly everything about the individual lives it observes while preserving enough structure to estimate broad national conditions. It can tell us whether unemployment is rising, whether labor force participation is falling, and whether employment is expanding. In July 2026 it reported an unemployment rate of 4.1 percent, a participation rate of 61.4 percent, and 6.9 million unemployed people — from about 39,000 conversations.

That is an extraordinary intellectual accomplishment. It is also a limited one.

03 A Compression Can Be Correct and Still Be Unintelligent

The three labor force categories were designed to answer a particular question:

**During the calendar week containing the twelfth of the month,
was this person working, looking for work, or neither?**

They were not designed to describe the complete economic structure of each person. And when a person holds more than one job, the survey resolves the ambiguity by rule: each employed person is counted exactly once, and their class of worker, industry, and occupation are assigned from **the single job at which they usually work the greatest number of hours**.

A surgeon earning \$900,000, a warehouse employee driving for a rideshare platform at night, and an AI consultant operating through a single-person corporation can all be classified as employed wage-and-salary workers. Their economic lives are completely different, but the model deliberately discards those differences.

The third case is not a rhetorical flourish. It is written into the definitions. BLS counts the *incorporated* self-employed as wage and salary workers, on the reasoning that a person who incorporates becomes, technically, a paid employee of their own corporation. Unless a table says otherwise, published self-employment figures cover only the unincorporated. So the founder of a one-person S-corp is, in the headline statistics, an employee.

That does not make the classification wrong. It makes the classification compressed.

The test is whether the discarded information becomes necessary for predicting what happens next. If the three categories continue to predict consumer spending, tax receipts, business formation, demand for labor, and economic resilience, then the compression remains intelligent. If those predictions begin separating from actual outcomes, the model has discarded something that now matters.

The failure appears first as surprise. Business applications surge while measured self-employment stays flat. Payroll growth weakens while economic production continues. Individuals create companies without leaving their jobs. New firms appear without hiring employees. Productive capacity expands without appearing as conventional labor force expansion.

The model calls these outcomes anomalies because it does not contain the structure necessary to anticipate them. **What looks like an anomaly in the data is often a limitation in the compression.**

FIGURE 02

Two instruments pointed at the same economy, disagreeing: year-over-year change, July 2025 to July 2026

Business applications, total		+23.4%
High-propensity applications		+5.9%
Applications from corporations		+3.9%
Self-employed, unincorporated		+1.2%
Self-employed, incorporated		-1.7%
Applications with planned wages		-20.0%

Americans filed 578,926 business applications in July 2026, up 23.4% from a year earlier and running roughly 60% above the 2019 pace on an annual basis. Over the same twelve months, applications that indicate planned wages — the ones expecting to hire — fell 20.0%, and the household survey's count of self-employed people barely moved. Something is being created that the labor force categories do not register. Sources: Census Bureau Business Formation Statistics, July 2026 (released August 12, 2026), seasonally adjusted; BLS Employment Situation, July 2026, series LNS12027714 and LNU02048984.

04 The Missing Category

The household survey was built for an economy in which a person was principally one of three things: an employee, a proprietor, or someone who was not working. The emerging economy contains another structure:

Employee + AI + business portfolio.

A person can remain employed by a bank while operating a consulting practice, a digital product company, and a specialized software business. AI supplies research, software development, marketing, bookkeeping assistance,

content production, and customer support. The person has not merely acquired a second job. **The person has become a small economic organization.**

The household survey still needs to classify that person once. If the bank consumes the most usual working hours, the person remains a wage-and-salary employee. The businesses can have employer identification numbers, customers, revenue, intellectual property, and growing economic value without changing the person's labor force classification. And if those businesses are incorporated, even the self-employment tables will not show them — the incorporated self-employed are counted as wage and salary workers by design.

The survey has correctly identified the person's main job. It has failed to preserve the person's economic architecture.

That distinction did not matter as much when starting and operating several businesses required employees, offices, vendors, and substantial capital. Those structures appeared elsewhere in the economy. They created payroll records, leases, bank loans, and visible organizations.

AI compresses the organization itself. One person can now contain economic capabilities that previously required an institution. The clearest statistical fingerprint of this is the divergence in Figure 02: business applications up 23.4 percent year over year, and applications that plan to pay wages down 20.0 percent. Entities are being formed that do not intend to hire anyone.

The old labor model compresses the person into “employee” at the same moment AI compresses an organization into the person. One compression is encountering another. The statistical category loses.

The survey is not blind to multiple work. It publishes a multiple jobholder count, and in July 2026 that count reached 8.7 million people, or 5.4 percent of the employed — a high for the recent series. But a multiple jobholder is still a person with jobs. It is not a person with a portfolio of entities, and the measure does not distinguish between the two.

05 Intelligence Is Not the Amount Compressed

The mathematical tradition behind minimum description length treats the best model, in its original two-part form, as the one that explains the data with the shortest total description: the model itself plus whatever remains unexplained by the model. Regularities permit compression. Randomness does not. A model earns its simplicity by accounting for what actually occurs.

A model that memorizes every observation predicts the past perfectly but compresses nothing. A model that reduces every observation to one constant compresses everything but predicts nothing. Intelligence exists between those failures. An intelligent system finds the smallest representation that preserves the structure required to predict unseen actuals.

That means compression ratio alone cannot measure intelligence. We must also measure what the compression can still do.

- 01 **Accuracy.** How closely does the prediction match what happens?
- 02 **Horizon.** How far into the future does the prediction remain accurate?
- 03 **Resolution.** At what level of detail does it stay right — the aggregate, the sector, the individual?
- 04 **Range.** Across how many different conditions does it hold?
- 05 **Calibration.** How well does it recognize when its confidence is unwarranted?
- 06 **Revision.** How quickly does it reorganize itself when actual outcomes reveal that its categories are wrong?

The superior intelligence does not necessarily retain more information. **It identifies more consequential information.**

06 Prediction Horizon

The most revealing of those measures is prediction horizon. Imagine two people examining the same situation. The first can predict what will happen during the next five hours. The second can predict what will happen during the next five years.

The second has found a deeper compression. The representation contains something more durable than the surface details of the present. It has captured structure that survives thousands of changing circumstances.

This does not mean every long-range prediction demonstrates superintelligence. An astronomical system can predict an eclipse decades ahead because celestial motion is highly regular. A human relationship can become unpredictable in an afternoon. Prediction horizon must therefore be judged within the same domain, at comparable resolution, and under comparable uncertainty.

The proper question is:

**How far can this compressed representation travel
before actual reality breaks it?**

Ordinary intelligence compresses enough structure to navigate the immediate future. **Exceptional intelligence** discovers patterns that remain valid after the visible circumstances change. **Superintelligence** identifies the small set of underlying relationships from which distant actuals can still be anticipated.

The distinction is not how much the system knows. It is how long its compressed understanding remains useful.

07 Actual Is the Judge

Intelligence cannot grade itself. A theory can be elegant. A model can be mathematically sophisticated. A report can be produced by the most respected institution in the country. None of that establishes intelligence. **Actual outcomes decide.**

A compressed model produces a prediction. Time converts the prediction into an actual. The difference between them reveals what the compression failed to preserve. Small differences mean the model retained the consequential structure. Large differences mean the model discarded information that mattered.

Repeated surprise is not bad luck. It is evidence that the representation is inadequate. This is worth holding against a real case: in the July 2026 report, the prior two months of payroll growth were revised down by a combined 103,000 jobs. Revision is the honest, visible form of a model correcting itself against actual. The question is whether the corrections are converging or accumulating in one direction.

This is also why prediction is a more useful test of intelligence than explanation. Human beings can construct compelling explanations after an event. Explanation can absorb every outcome because it is permitted to use the outcome itself. Prediction must commit before the actual arrives.

Prediction places the compression at risk.

08 Judging the Household Survey

The household survey is highly intelligent within its original purpose. It converts a vast population into a manageable representation and produces useful estimates of national employment conditions within weeks. BLS itself is candid about the strain: estimates have remained reliable despite falling response, but a continued decline would slowly erode the survey's ability to detect meaningful change. A web self-response mode is planned for introduction in 2027.

Its intelligence declines when we ask it to see economic structures it was not designed to preserve.

Question put to the survey	Series	What it preserves	Horizon	What it discards, and why that now matters
Did this person work last week?	CPS Table A-1	Employment status for the reference week, weighted to the national population.	STRONG	Almost nothing consequential. This is the question the instrument was built for, and it still answers it well.
How many people are out of work?	CPS Table A-1	Active job search in the last four weeks plus current availability.	STRONG	People who stopped searching. The published measure is supplemented by marginally attached and discouraged worker counts — 1.8 million and 476,000 in July 2026 — which exist precisely because the headline category is known to be lossy.
What kind of worker is this person?	CPS Table A-8	A single class of worker, taken from the job with the greatest usual hours.	WEAK	Every other economic role the person holds. A one-person incorporated business is recorded as wage and salary employment, which means the fastest-growing form of solo enterprise is invisible in the self-employment series by construction.
Is entrepreneurship rising?	CPS vs. Census BFS	A count of unincorporated self-employed people, essentially flat year over year.	WEAK	Entity formation. Business applications rose 23.4% over the same twelve months. The two instruments now tell different stories, and only one of them is designed to see firms that never hire.

Question put to the survey	Series	What it preserves	Horizon	What it discards, and why that now matters
Is productive capacity expanding?	CPS + CES payrolls	Headcount — people employed, jobs on payrolls, hours worked.	WEAK	Capability per person. When one worker plus AI performs work that previously required a team, output can rise while every headcount series falls. The model has no field for leverage.

Fidelity of the Current Population Survey by question asked. Sources as listed under Figures 01 and 02, plus BLS Tables A-1, A-8 and the CPS concepts and definitions page.

The survey can tell us whether someone worked. It is less capable of telling us what that person has become. It can classify a main job. It cannot adequately represent a portfolio of simultaneous economic roles. It can count unincorporated self-employment. It struggles to see an employee using AI to operate several businesses without employees.

The problem is not that 39,000 is automatically too small. The problem is that the compression removes distinctions that are becoming causally important. Sample size and category design are separate failures, and only one of them is fixed by knocking on more doors.

If business applications, production, income, and entrepreneurial activity increasingly diverge from the three labor force categories, the model's predictive horizon contracts. It can still describe the reference week. It becomes progressively less capable of describing the economy being formed.

That is how we judge its intelligence. Not by its institutional authority. Not by the size of its dataset. Not by the elegance of its methodology. By the distance between what its compressed categories predict and what subsequently becomes actual.

09 The Measure of Intelligence

Intelligence is the ability to discard almost everything and still be right.

The greater the compression, the more impressive the intelligence — but only while predictive accuracy survives. Once the prediction separates from actual, additional compression is no longer intelligence. It is information loss.

A regular intelligence sees enough pattern to predict five hours. A superior intelligence finds structure that remains predictive for five years. A superintelligence does not carry a larger pile of facts into the future. It carries a smaller and more accurate representation of what matters.

That gives us a practical way to judge intelligence in a person, an institution, or an artificial system.

- 01 Give it reality.** Hand the system an unedited situation, not a curated one. The compression has to choose what to throw away without being told what the answer will be.
- 02 Observe how it compresses.** Watch what it keeps and what it discards. The categories it invents are the theory it is actually holding, whatever it says its theory is.
- 03 Ask it what comes next.** Require a commitment, at a stated resolution and horizon, before the outcome exists. Explanation is free; prediction is not.

04 Then wait for actual. Measure the distance. Small distance means the compression kept what mattered. Large or one-sided distance means the categories need rebuilding, not more data.

The intelligent model is not the one that says the least. It is the one whose few words continue to be right.

Sources

- U.S. Bureau of Labor Statistics. *The Employment Situation — July 2026*. USDL-26-1291, released August 7, 2026. bls.gov/news.release/empsit.nr0.htm
- U.S. Bureau of Labor Statistics. *Employment Situation Technical Note*. bls.gov/news.release/empsit.tn.htm
- U.S. Bureau of Labor Statistics. *Employment Situation Summary Table A* (household data, seasonally adjusted). bls.gov/news.release/empsit.a.htm
- U.S. Bureau of Labor Statistics. *Table A-8. Employed persons by class of worker and part-time status*. Footnote 1: wage and salary workers include the incorporated self-employed. bls.gov/news.release/empsit.t08.htm
- U.S. Bureau of Labor Statistics. *Table A-16. Persons not in the labor force and multiple jobholders*. bls.gov/news.release/empsit.t16.htm
- U.S. Bureau of Labor Statistics. *Labor Force Statistics from the Current Population Survey: Concepts and Definitions*. bls.gov/cps/definitions.htm
- U.S. Bureau of Labor Statistics. *Current Population Survey Response Rates*. bls.gov/cps/methods/response_rates.htm
- U.S. Bureau of Labor Statistics. *CPS overall response rate, series LNU09300000*. July 2026 = 65.3 percent. data.bls.gov/timeseries/LNU09300000
- U.S. Bureau of Labor Statistics. *Handbook of Methods: Current Population Survey, Design*. bls.gov/opub/hom/cps/design.htm
- U.S. Census Bureau. *Current Population Survey: Methodology*. census.gov/programs-surveys/cps/technical-documentation/methodology.html
- U.S. Census Bureau. *Business Formation Statistics, July 2026*. Release CB26-130, August 12, 2026. census.gov/econ/bfs/pdf/bfs_current.pdf
- Rissanen, J. *Modeling by shortest data description*. *Automatica* 14 (1978): 465–471.
- Grünwald, P. *The Minimum Description Length Principle*. MIT Press, 2007.