Why does a number you have never seen decide whether you get a loan?

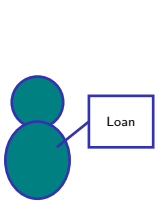
A single number determines your access to housing, education, and economic opportunity.

The power of the score:

- Compresses your entire financial life into one number
- Updated constantly without your knowledge
- Used by lenders, landlords, employers
- Errors can lock you out for years

The opacity:

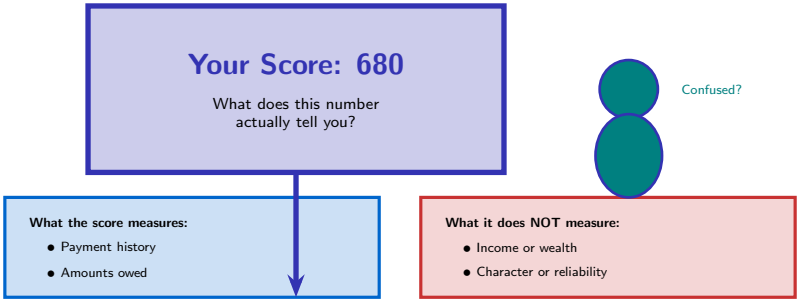
- Algorithm is proprietary and secret
- You cannot see how it works
- Appeals process is limited
- Wrong decisions have no accountability



Credit scores are gatekeepers to opportunity—yet most people do not understand how they work or who controls them.

A number you never see, calculated by algorithms you cannot audit, determines your life chances.

Do you know your credit score—and do you know what it actually measures?



Most people conflate credit scores with creditworthiness—they are not the same.

Reflect: Do you know your score? Do you know what behaviors improve or harm it? Most people do not—yet it shapes their lives.

Credit scores measure past borrowing behavior, not trustworthiness or future financial capacity.

What are the different approaches to scoring creditworthiness?

Traditional credit bureau scoring:

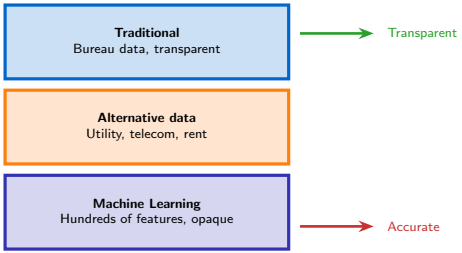
- Uses payment history, debt levels, length of history
- Requires existing credit file
- Transparent methodology, decades of validation
- Excludes those without formal credit history

Alternative data scoring:

- Uses utility payments, rent, telecom data
- Works for thin-file or no-file borrowers
- Less validated, more privacy concerns
- Enables access but with accuracy trade-offs

Machine learning scoring:

- Uses hundreds of features automatically
- Higher accuracy, captures non-linear patterns
- Opaque, difficult to explain or audit
- Raises fairness and bias questions



Each approach makes different trade-offs.

No scoring approach is perfect—each balances transparency, accuracy, and access differently.

The choice of scoring approach determines who gets access and on what terms.

How does a credit scoring model turn your financial history into a single number?

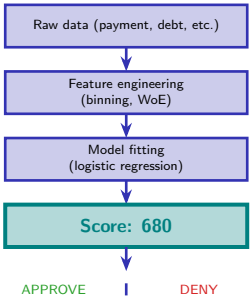
The scoring pipeline:

1. Collect raw data (payment history, balances, inquiries)
2. Transform features (binning, encoding)
3. Calculate feature weights based on predictive power
4. Fit statistical model to predict default probability
5. Convert probability to score (higher is better or worse)
6. Set cutoff threshold for approve versus deny

The compression:

- Hundreds of data points → one number
- Complex financial life → binary decision
- Nuance lost in aggregation

The problem: Every compression discards information—what gets lost matters for fairness.



Complex life reduced to a single number.

The scoring pipeline is a compression algorithm—and compression always loses information.

Understanding the pipeline reveals where bias can enter and where fairness can be compromised.

How do traditional and alternative credit scoring architectures compare?

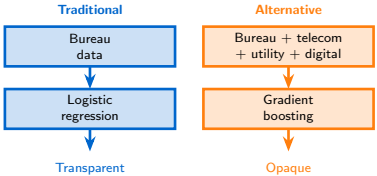
Traditional scoring architecture:

- Data source: Credit bureau reports
- Features: 15–30 hand-engineered variables
- Model: Logistic regression
- Output: Probability of default
- Validation: Decades of back-testing
- Regulation: Well-understood, auditable

Alternative scoring architecture:

- Data source: Bureau plus telecom, utility, digital footprint
- Features: 200–2000 auto-generated variables
- Model: Gradient boosting, neural networks
- Output: Probability of default
- Validation: Limited historical data
- Regulation: Uncertain, evolving

The core trade-off: Traditional is transparent but exclusive; alternative is inclusive but opaque.



Architecture determines transparency and access.

The shift from traditional to alternative scoring trades interpretability for inclusion—neither is clearly superior.

Architectural choices encode values: who gets access, who can audit, who bears the risk of errors.

What goes wrong when a credit model learns from biased historical data?

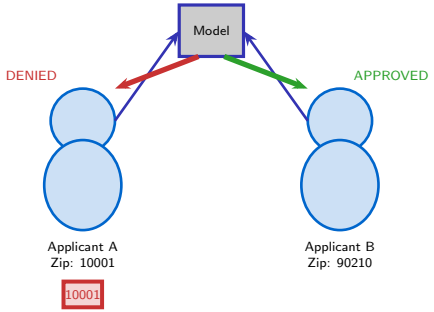
The bias feedback loop:

1. Historical data reflects past discrimination
2. Model learns patterns from biased data
3. Model predicts denial for historically excluded groups
4. Denied applicants generate no repayment data
5. Missing data reinforces original bias
6. Cycle repeats indefinitely

Proxy discrimination:

- Zip code correlates with race
- University attended correlates with class
- Phone type correlates with income
- Model never sees protected characteristics directly
- But proxies encode the same information

The result: Algorithmic discrimination looks neutral but perpetuates historical inequities.



Proxies encode discrimination without using protected characteristics directly.

Models trained on biased data perpetuate bias—even when protected characteristics are excluded from the model.

Historical bias becomes algorithmic bias; removing protected variables is necessary but not sufficient for fairness.

Where do alternative data sources improve scoring accuracy most?

Alternative data value by segment:

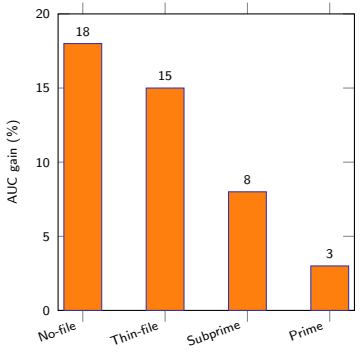
- Thin-file borrowers: Largest accuracy gain
- No-file borrowers: Enables first-time scoring
- Prime borrowers: Minimal improvement
- Subprime with history: Moderate improvement

Which data sources help most:

- Utility payment history (electricity, water)
- Rent payment history (often unreported)
- Telecom payment patterns (mobile top-ups)
- Transaction velocity and regularity
- Digital footprint consistency

Geographic variation:

- Developing markets: Mobile data dominates
- Developed markets: Utility and rent data



Alternative data helps most where traditional data is absent.

Alternative data creates the largest accuracy improvements for those historically excluded from credit markets.

The benefit of alternative data is not uniform—it is concentrated among the previously unscoreable.

Who benefits and who is harmed by the current credit scoring system?

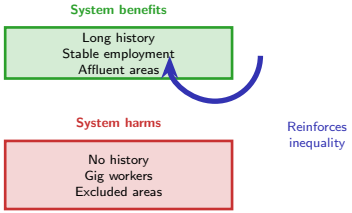
Who benefits:

- Prime borrowers with long credit histories
- Those with stable employment and assets
- Homeowners and degree holders
- Residents of affluent zip codes
- Dominant demographic groups

Who is harmed:

- Young adults with no credit history
- Immigrants and displaced populations
- Gig workers and informal earners
- Residents of historically redlined areas
- Anyone who opted out of credit markets

The system property: Credit scoring rewards those who already have access and penalizes those who were excluded.



The system amplifies existing advantages.

Credit scoring is not neutral—it systematically advantages those with prior access and disadvantages those without.

The current system treats past exclusion as evidence of risk, creating a self-fulfilling prophecy.

Four questions to evaluate whether a credit scoring model is fair

The Fair Scoring Audit:

Question 1: Does the model use protected characteristics as proxies?

- Check correlations between features and race, gender, age
- Audit zip code, university, employer for proxy effects
- Test for disparate impact across groups

Question 2: Are approval rates equitable across groups?

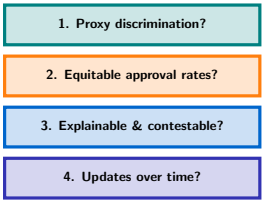
- Compare approval rates for equivalent risk levels
- Measure false negative rates by demographic
- Test calibration separately for each group

Question 3: Can the applicant understand and contest the decision?

- Provide specific reasons for denial
- Offer clear path to improve score
- Enable appeals with human review

Question 4: Does the model update as society changes?

- Monitor for population drift
- Retrain on recent data
- Adjust for macro shocks (recession, pandemic)



Apply all four tests rigorously.

Fair scoring requires passing all four tests—proxy avoidance, equitable outcomes, explainability, and adaptability.

Your Challenge

Design Challenge: Alternative Scoring Model

Your dataset:

- Loan applications from people with no credit bureau file
- Available data: mobile top-ups, utility payments, transaction patterns
- Demographics: age, location, education, employment type

What do you include, exclude, and why?

- Which features do you use?
- Which do you exclude to avoid proxy discrimination?
- How do you test for bias?
- How do you explain denials to applicants?

Build: Design the feature set and justify each choice
Audit: Apply the four-question fairness test
Defend: Would you deploy this model? Why or why not?

This challenge forces confrontation with the accuracy-fairness-explainability trade-offs inherent in credit scoring.

Real credit scoring requires navigating trade-offs with no perfect solution—your choices reveal your values.