

PAPER · OCTOBER 2026

From Stamps to Evidence

How Humans Learned to Trust Strangers, and What Changes
When Agents Do the Checking

FIROZ AZEES, ILAN000

ILAN000.COM/FROM-STAMPS-TO-EVIDENCE

CONTENTS

- | | |
|----|--|
| 01 | Trust is the oldest infrastructure |
| 02 | Four eras of trusting strangers |
| 03 | The law underneath every trust signal |
| 04 | What AI broke |
| 05 | The transition: when signals are cheap, choice falls back on what it already knows |
| 06 | The other half: checking became cheap too |
| 07 | What agents will weigh |
| 08 | Brands, students and professionals: three different problems |
| 09 | Democracy and scoring: who computes trust |
| 10 | Mobility: the opening and the lock-in |
| 11 | Seven principles for a trust infrastructure worth having |
| 12 | What we do not know |
| 13 | Where this leaves us |

SOURCES

References

Summary

The argument in one line. Every system humans made to trust strangers was a way of making checking cheaper. Institutions issued stamps because checking a person was expensive. AI has made two things cheap at once: making anything, and checking anything. The first broke the stamps. The second can replace them with evidence, read by agents that work for each of us, if we build it that way.

How we got here. Trust moved through four eras. In the village we watched each other, which could not be faked and could not grow past a few hundred people. Reputation networks (guilds, merchant coalitions, apprenticeships) carried trust between towns. Institutions issued stamps (degrees, licences, brands, credit scores), so a stranger could be trusted without being watched. Platforms added ratings from the crowd. Each era solved the scale problem of the one before, and each moved the person who issues the signal further from the person who pays when it is wrong. (Chapter 2)

The law underneath. A signal can be trusted only while faking it costs more than having the real thing, and only while whoever issues it answers for it. (Chapter 3)

What broke. AI made the essay, the cover letter, the review, the portfolio and the polished page nearly free to produce. The early evidence is in: on one large freelancing platform the link between a tailored cover letter and a callback fell by half, and at one UK university almost every AI-written exam answer went unnoticed. (Chapter 4)

What happens first. While signals are cheap and checking is not yet built, the early evidence suggests choosers fall back on what they already know: the incumbent brand, the familiar school, the friend's referral. This is the freezing the agents essay warned about. (Chapter 5)

What comes next. Checking is getting cheap too, though not yet reliable. An agent can read hundreds of sources, compare them, remember how its last choice turned out and ask the other side's agent for proof. The infrastructure for that is being built now. Trust can move from stamps to evidence, and from institutions to the agents that check for us. (Chapters 6 and 7)

Three different problems. Brands leave dense public evidence and will be read first. People leave little, most of it private, and most people will never publish. For them evidence will come as a by-product of working with AI, and institutions will change from issuers of stamps into witnesses of what happened. (Chapter 8)

What decides whether it is fair. Who computes trust (an authority that scores everyone, or each person's own agent), who owns the record of a person's work, and whether newcomers get a real trial. (Chapters 9 and 10)

What we propose. Seven principles for a trust infrastructure worth having. (Chapter 11)

What we do not know. A great deal, listed honestly. Much of the agent evidence is from simulations and preprints. (Chapter 12)

Trust is the oldest infrastructure

Groups survive. Individuals die. That has been true for as long as there have been humans, and it is the starting point of everything I write.

A lone human is slow, has no claws, sees badly at night and needs years of helpless childhood before being any use. Cooperation is not a strategy for such an animal. It is the minimum requirement to exist. But cooperation has a precondition that is easy to miss because it is everywhere: before I hand you my grain, my child, my savings or my vote, I have to believe you will not run with it. Every bigger thing humans have done (cities, trade, science, the company, the state) rests on some way of answering one question about a stranger. Can I trust this person?

We do not usually call the answer infrastructure, because it is made of habits, papers and reputations rather than roads and pipes. It is infrastructure all the same. It has a cost to build and maintain, it decides who gets in and who stays out, and when it fails, cooperation fails with it.

This paper is about that infrastructure: how it was made, why it is breaking now, and what replaces it. I come to it from a particular place. For most of my life I watched the systems we made to sort people (exams, degrees, interviews, the right accent, the right name on a CV) and saw that they were measuring something other than what they claimed. They measured how much a person's environment had let them show. Then they called the result ability. That is a category error, and it decides who gets trusted with opportunity.

AI has now put pressure on every one of those systems at once. The essay that proved a student could think, the cover letter that proved an applicant cared, the review that proved a restaurant was good, the polished website that proved a company was real: each of these can now be produced in seconds by anyone. When the proof becomes free, it stops being proof.

The usual reaction is alarm. My argument is different. The same technology that broke our trust signals has also made checking cheap, and checking is what trust was always made of. Whether the next trust infrastructure is fairer than the last depends on choices being made right now, mostly by people who are not thinking about fairness at all.

Four eras of trusting strangers

2.1 The village: trust by watching

In a small group, trust is simple. You claim to be a good tracker; everyone watches you track. If you fail, you go hungry, and so does everyone who believed you. In the essay *The Collapse of Talent Signals* I described the three properties this gave the earliest human groups: everyone saw the same evidence at the same time, a claim was tested within days, and the people who judged a claim lived with the result. [1]

Village trust still works this way. In a small town, people know whose family you are from, who your grandfather was, what you did at fourteen. A bad act is checked by gossip before any court hears of it. The sociologist Robert Sampson and his colleagues called the neighbourhood version of this *collective efficacy*: trust between neighbours together with their willingness to step in. Across 343 Chicago neighbourhoods it explained much of the link between disadvantage and violence. [2]

The village has one virtue that every later system lost. You could not fake it. You were watched for years by people who had to live with you.

It has one limit that every later system tried to escape. It does not scale. The anthropologist Robin Dunbar estimated, from the brain sizes of primates, that humans can keep around 150 stable relationships. [3] The exact number has been argued over since. The point survives the argument: a group can only be watched by its own members up to a few hundred people, and beyond that trust by watching runs out.

2.2 Reputation networks: trust carried between towns

When trade moved beyond the village, trust had to travel. It travelled through networks. A guild master vouched for a journeyman after years of watching his work. A merchant in one port trusted an agent in another because both belonged to the same community, and cheating one member meant losing the whole network's business.

The economist Avner Greif made the most famous version of this argument, about eleventh-century Maghribi traders in the Mediterranean who, he argued, enforced honesty through a coalition that shared information about cheats. [4] Later historians have disputed how much of that coalition the documents actually show. [5] Edwards and Ogilvie's reading is that the traders also used courts and contracts. The general point survives the dispute: before institutions could reach far enough, much of the trust between strangers was carried by people who knew both sides and had something to lose.

Two things changed in this era. For the first time the signal (a guild's certificate, a merchant's name) was separated from the behaviour it stood for. And for the first time the signal could be bought, inherited or kept inside a closed group. Guilds restricted

entry. Families passed down names that the next generation had not earned.

2.3 Institutions: trust by stamp

The industrial economy needed something networks could not give: a way to trust millions of strangers who would never share a guild or a port. The answer was the stamp.

A school certificate, a university degree, a professional licence, a brand, a credit score. Each one is a promise, made by an institution, that someone has already done the checking. The sociologist Lynne Zucker described this shift in 1986. She separated three ways trust gets produced: through a history of exchanges between two parties, through shared characteristics such as family or ethnicity, and through institutions. In the America she studied, the third grew because the first two could not keep up with a mobile, industrial society. [6]

The stamp solved a real problem. Checking a person properly was expensive. An employer could not watch every applicant work for three years. A lender could not interview every borrower's neighbours. A parent could not inspect every school. So we paid institutions to check once, issue a stamp, and let the stamp travel.

The credit score is the cleanest example of how this works when it works. Fair, Isaac and Company was founded in 1956 and the first FICO score arrived in 1989; before that, the US Fair Credit Reporting Act of 1970 required the bureaus to follow reasonable procedures so that what they reported was fair and accurate. [7] Several bureaus compete, the data has a standard shape, and a person can challenge an error. It is far from perfect, but it is governed.

Most stamps were not governed like that, and they carried a cost the credit score's history makes easy to miss. A stamp only reports what the institution could see, and what an institution can see is what a person's environment let them show. Fees, coaching, the right school district, confidence that comes from never having been told you are average. The filter calibrated itself to environmental expression and then called the result merit. Michael Young saw this coming in 1958, in *The Rise of the Meritocracy*, a satire about a society that sorts people by measured intelligence plus effort and hardens into a new aristocracy. [8] Young meant the book as a warning. We took the word he made famous and used it as a goal.

The sociologist Lauren Rivera later showed the mechanism at work inside elite American firms: hiring that claimed to screen for ability screened heavily for class-coded signals of fit, the same leisure activities, the same polish. [9] The stamp made strangers trustable at scale. It also made privilege look like merit.

2.4 Platforms: trust by crowd

The internet promised to fix the stamp. Anyone could publish; anyone could be found. Rachel Botsman named the next era *distributed trust*: trust that flows sideways, to platforms and to the ratings of strangers, rather than upward to institutions. [10] You get into a stranger's car because four hundred other strangers gave the driver five stars.

The platforms solved discovery. They did not solve truth. Ratings have two well-known problems. The first is who writes them. Only the very pleased and the very angry bother, so the scores pile up at five stars and one star, and the average describes the reviewers more than the product. Nan Hu, Jie Zhang and Paul Pavlou described this shape, the J-curve, in 2009. [11] The second is that ratings drift upward over time. Apostolos Filippas, John Horton and Joseph Golden showed why: raters feel pressure not to hurt the person they rate, so scores creep up until they stop separating good from bad. They called it reputation inflation. [12]

For courses and colleges the problem is worse. A meal is judged when you finish it. Education pays off, or does not, over years. The review is collected at the end of the course, before anyone knows whether it worked, so it rewards the course that felt pleasant rather than the one that changed someone's working life. A five-star system built for restaurants was being asked to judge something whose result arrives after the reviewer has gone.

Ratings still move behaviour. Michael Anderson and Jeremy Magruder found that an extra half star on Yelp made a restaurant about 19 percentage points more likely to sell out its tables. [13] That is exactly why they were worth gaming, and why they were gamed.

2.5 The pattern across the four eras

Put the four eras side by side and one pattern stands out.

ERA	HOW A STRANGER IS TRUSTED	WHAT IT SOLVED	WHAT IT LOST
Village	Watching, over years	Could not be faked	Runs out beyond a few hundred people
Reputation networks	Vouching by people who know both sides	Trust between towns	Closed groups, inherited names
Institutions	A stamp issued once, carried everywhere	Trust between millions	Stamps measure access, not ability
Platforms	Ratings from many strangers	Discovery at global scale	Inflated, easily gamed

Each era solved the scale problem of the one before. And each moved the person who creates the signal further from the person who pays when it is wrong. In the village they were the same people. In the guild they knew each other. The university keeps its fees whether or not its graduate can do the job. The platform earns from the review whether or not the review is true.

The law underneath every trust signal

Two ideas explain why some signals hold and others collapse.

The first comes from economics. In 1973 Michael Spence asked why employers trust degrees at all. His answer was that a signal works only when it costs more for the wrong kind of person to send it than for the right kind. [14] A degree said something because finishing one was harder for someone who could not do the work. Biology found the same law from the other side. The biologist Amotz Zahavi argued in 1975 that animals trust signals that are expensive to make (the peacock's tail is believable because a weak bird cannot carry it). [15] Zahavi's idea is still argued over by biologists. Alan Grafen gave it a formal footing in 1990, in the same terms as the economists: a signal is honest when faking it costs more than it is worth. [16]

The second idea is mine, and I have written it before. I call it *consequence separation*. When the person who issues a signal, the person who uses it and the person who suffers when it is wrong are the same, signals stay accurate. Separate them and the signal degrades. Separate them completely and you get what we have now: a university that is paid whether its graduate succeeds or not, a review site that profits from every review, a reference letter written by someone who will never work with the person again. [1]

George Akerlof described the end state in 1970. In a market where buyers cannot tell good from bad, the bad drives out the good, because the seller of the good car cannot prove it is good. [17] He called it the market for lemons. A trust infrastructure is what stops a society from becoming one.

Put the two ideas together and you have a test any trust signal can be put through:

1. **Is it expensive to fake, relative to having the real thing?**
2. **Does whoever issues it pay when it is wrong?**

The village passed both. The stamp passed the first for a long time and the second rarely. The rating passed neither for long. Keep this test in mind, because the rest of the paper uses it.

What AI broke

AI did not invent the weaknesses of our trust signals. It removed the one thing that kept most of them standing: the cost of producing them.

An essay used to say that its writer could think, because writing a good one took thinking. A cover letter tailored to one company said that the applicant cared, because tailoring took time. A detailed review said that someone had really eaten the meal. A polished website said that a company had money and people behind it. Each of these passed Spence's test only because producing it was hard.

Now each costs almost nothing. The evidence that this has already broken the signals is no longer speculation.

On a large online freelancing platform, Anaïs Galdin and Jesse Silbert compared the period before and after large language models arrived. Before, employers paid a premium for applications written for the specific job. After, the premium disappeared. Their model of the market suggests that when written applications carry no signal, the best workers get hired about a fifth less often and the weakest more often, because employers can no longer tell them apart from what they wrote. [18] When the same platform launched its own AI cover-letter tool, Jingyi Cui, Gabriel Dias and Justin Ye found that the link between how closely a letter was tailored to the job and whether the applicant got a callback fell by about half (51 per cent), and employers leaned on other signals instead. [19] Both are working papers, not yet peer reviewed. They point the same way.

In universities, Peter Scarfe and colleagues slipped exam answers written entirely by AI into five real undergraduate modules at a UK university. Ninety-four per cent went undetected, and on average they scored half a grade boundary higher than the students' own work. [20] An unsupervised essay can no longer tell a marker who did the thinking.

In reviews, the early evidence points the same way. In one small 2025 study, people asked to tell AI-written product reviews from human ones did about as well as a coin toss, and seven language models given the same task did no better. [21] A machine reading reviews is not yet a better judge of them than we are.

There is a twist that makes this worse. When employers use AI to screen applications that applicants wrote with AI, the screening models favour text that sounds like their own. Jiannan Xu, Gujie Li and Jane Yi Jiang found that screening models ranked résumés they had written themselves above human-written ones, and that applicants who used the screener's own model were more likely to be shortlisted. [22] The signal does not just weaken. It starts to reward whoever happened to use the right tool.

None of this means people stopped being capable. It means the things we used to read for capability stopped carrying it. The cost of faking fell to zero, so by Spence's law the signal had to die.

The transition: when signals are cheap, choice falls back on what it already knows

When the old signals stop working and nothing has replaced them yet, people do not stop choosing. They fall back on what they already know.

Employers show the first signs of it. When written applications stopped telling candidates apart, Cui and colleagues found that employers leaned on other signals instead. [19] When large American companies dropped degree requirements from their job ads, the Burning Glass Institute and Harvard Business School researchers found that actual hiring barely moved. Where a requirement was dropped, the share of hires without a degree rose by about three and a half percentage points, but across all hiring that came to fewer than one in 700 hires, and almost all of the change came from about a third of the firms. [23] Removing a filter did not remove the habit of filtering. My expectation, which no study has yet measured, is that managers who lose one filter reach for an older one: the familiar company name on a CV, the friend who vouches, the person who feels like a fit.

Referrals are the purest form of this fallback, and they carry real information. Amanda Pallais and Emily Sands showed that referred workers on an online platform performed better, which is why employers trust them. [24] But a referral reaches only people who already know someone. When trust narrows to who you know, the person with no network becomes invisible again.

Whole societies show the same reflex. Edelman titled its 2026 Trust Barometer *Trust Amid Insularity*: seven in ten people across 28 countries say they are unwilling or hesitant to trust someone with different values, facts or background. [25] The institution that issued the most important stamp of all is losing ground too: confidence in American higher education, as Gallup and the Lumina Foundation measure it, slipped to 38 per cent in 2026, against 57 per cent in 2015. [26]

AI agents, which are now starting to choose products and services for people, show the same pull. In a 2025 experiment in a simulated marketplace, Amine Allouah and colleagues found that AI shopping agents were swayed by where an item sat on the page and by platform endorsements, and penalised items marked as sponsored. [27] In a 2026 preprint on product recommendations, Xi Chu and Yupeng Hou found that a well-known skincare brand beat rivals with identical specifications every time, though a small edge in ratings could break its lead. [28] Earlier work found that language models link global brands with positive traits more readily than local ones. [29] In that study, when the evidence was equal, the familiar name won.

In my essay *Agents Remember* I described what happens next if nothing changes. An agent that booked the same café three times stops searching. An agent that trusted one creator keeps recommending her. Multiply that by a billion agents and you have a memory with a power law built in: whoever gets in early is hard to remove. [30] Human institutions froze hierarchy because they remembered. Agents can freeze it again, faster.

This is the transition, and it is where we are now. It is not where we have to stay.

The other half: checking became cheap too

Everything in chapter 4 is about one effect of AI: making things became cheap. There is a second effect, and it matters more for trust.

Checking became cheap.

Go back to why the stamp existed. Institutions issued stamps because checking a person was expensive. An employer could not read every applicant's work, call three former clients and test them on a real problem. A buyer could not read every review, compare every price, and find out how the last hundred customers felt six months later. A parent could not visit every school. So we paid someone to check once and we trusted their stamp.

Checking one job applicant properly once meant reading their work, phoning the people they had worked for and testing them on a real problem: days of a manager's time, which is exactly why most firms read the degree instead. An agent can do much of that checking. It can read hundreds of sources in the time a person reads one. It can compare what a company claims with what independent sources say. It can ask the other side's agent for its records. It can remember how its last choice for you turned out. A structured interview that once needed a trained person's afternoon can now be run by software for thousands of applicants at once (chapter 8 describes one study). The due diligence that only rich buyers, admissions offices and large employers could afford is becoming available to anyone with an agent.

Cheaper is not the same as reliable. The same evidence that shows checking getting cheap shows machine checkers have biases of their own: screening models that favour their own writing, shopping agents swayed by where an item sits on the page. [22] [27] Agents can check. Whether they check fairly is a design question, not a given.

If they are built well, this is the next era of trust:

ERA	TRUST COMES FROM	WHY IT WORKED	WHY IT BROKE
Village	Watching	Could not be faked	Did not scale
Reputation networks	Vouching	Carried trust between towns	Closed groups
Institutional	Stamps	Checking was expensive, so we paid someone to check once	Stamps measured access more than ability
Platform	Ratings	The crowd at scale	Inflated, and now free to fake

ERA	TRUST COMES FROM	WHY IT WORKED	WHY IT BROKE
Verified	Evidence, checked by your own agent	Checking became cheap	(not yet known)

In one sense this era is the first one coming back. In the village you watched the person yourself. Now your agent watches the evidence for you, with no village-sized ceiling and no institution standing in between. The internet opened information to anyone with a connection, and AI is opening up intelligence. Verification, the due diligence only institutions could afford, is the next thing to open.

The machinery for this is already being built, and it is mostly not about trust in the abstract. It is about payments and identity, because that is where the money is.

- **Agents talking to agents.** An open protocol that lets agents from different companies describe what they can do and pass tasks between them was released in April 2025 and is now run under the Linux Foundation. [31]
- **Telling a trusted agent from a bot.** The card networks are building this first. Visa's protocol lets an agent pass signed information about its intent and the shopper it acts for, so a merchant can tell a trusted agent from a bot. Mastercard says agents on its network will be registered and verified, and that shoppers can set limits on what their agent may buy. [32]
- **Knowing which agent you are dealing with.** Researchers studying the infrastructure agents will need, including Alan Chan and colleagues, argue for ways to identify an agent, to tie it to the person or company responsible for it, and to keep records that can be checked when something goes wrong. [33]
- **Credentials a person carries.** The World Wide Web Consortium published Verifiable Credentials 2.0 as a standard on 15 May 2025: a format for a signed claim (a degree, a licence, a membership) that the holder keeps and shows when asked. [34] Under the revised eIDAS regulation, the European Commission says member states must offer citizens a digital identity wallet by the end of 2026. [35] India's DigiLocker already stores, shares and verifies government-issued documents, and APAAR, a permanent twelve-digit academic identity for students, shows more than 33 crore registrations. [36]
- **Where content came from.** The C2PA specification defines content credentials: a signed record of where a piece of media came from and how it was changed. [37] Its limit is built in. A signature can say where something came from. It cannot say whether it is true.

Notice who is building most of this. Card networks, standards bodies and governments. The trust layer for agents is not turning out to be a free-floating crowd. It is institutions, again, but doing a different job. I return to this in chapter 8.

What agents will weigh

If an agent is going to decide whom to trust on your behalf, what will it look at? Go back to the test from chapter 3. A signal holds when it is expensive to fake and when someone answers for it. An agent that is any good will weigh exactly the things that still pass that test when producing anything is free.

Time. A track record built over years cannot be generated overnight. A long record resists faking, because faking it would take as long as doing it. But time is also the incumbent's advantage. An agent that weighs only how long someone has been around will lock out every newcomer, which is the freezing of chapter 5. A fair agent weighs how recent the record is and how fast it is improving, not only how long it is (chapter 10).

Independent agreement. One glowing source is cheap. Many sources that agree, and that do not share an owner, are expensive to fake: a single page can be manufactured in a minute, but agreement across independent sources cannot easily be. The hard part for an agent is telling real independence from a network of sites pretending to be many.

Outcomes. Did the buyer come back? Did the hire stay a year? Did the student's project get used by anyone? Did the course lead to the job? This is where agents will change the most. An agent remembers what it chose and how it turned out. Across millions of agents, that memory of outcomes will replace the five-star review. The review was a human typing an opinion on the day of the purchase. The outcome record is what actually happened, months later. It answers the problem I saw in course reviews: the result arrives after the reviewer has gone. An agent does not leave.

Accountability. Who answers if this is wrong? A claim signed by a company that can be held to it is worth more than the same claim from nobody. This is why so much of the new infrastructure in chapter 6 is about tying an agent to a responsible owner.

Process. How was this made? For work done with AI, the finished artifact proves little, but the record of decisions (what was asked, what the machine gave, what the person kept, changed or rejected) shows who did the thinking. Ilanooo's Direction Report reads a child's build this way, from her decisions rather than from the polish of the result. [38]

Stake. Who loses if the claim is false? A guarantee, a refund, a reference from someone whose own reputation suffers if the person fails. This is consequence separation run backwards: put the cost back on whoever issues the signal, and the signal becomes honest again.

The agents that weigh these things will not be neutral, and they will not all agree. Each will be tuned to the person it works for. One person's agent will care most about price and refunds; another's about whether a company treats its workers well; another's about whether an opinion it reads was paid for. That variety is not a flaw. It is what keeps the next trust infrastructure from becoming one score that everyone games.

Brands, students and professionals: three different problems

The next era will not arrive for everyone at once. Brands, students and working professionals present agents with three very different kinds of evidence.

8.1 Brands will be read first

A brand leaves a dense public trail: thousands of buyers, transactions, prices, reviews, returns, mentions in the press and in forums. Money flows through it, so the payment infrastructure in chapter 6 is being built around it first. An agent can cross-check a brand's claims against what independent sources say, and it will. The brands with thin evidence will be found out first.

8.2 People are harder, for four reasons

There is little evidence. A brand has thousands of buyers. A professional has perhaps twenty or fifty clients and colleagues. A student has a handful of teachers.

Most of it is private. Work happens inside companies, under confidentiality agreements, inside classrooms. An outside agent cannot see it.

Most people will not publish, even when producing is free. For two decades, a small share of people online have made most of what everyone else reads. Jakob Nielsen named the rule of thumb in 2006: in most online communities about 90 per cent only read, 9 per cent contribute now and then, and 1 per cent make most of what there is. [39] AI lowers the cost of producing, but not the will, the confidence or the time to put oneself in public. And public output is now the cheapest signal there is, so a flood of it proves nothing. A trust system built on publishing would reward the confident and the time-rich: a new privilege dressed as openness.

It is regulated, and should be. The European Union's AI Act places AI systems used in education and in employment decisions in its high-risk category, with obligations on whoever builds and deploys them. [40] Evidence about people will move slower than evidence about products.

8.3 What the people side becomes

Four changes seem likely to me, and they are different from what the brand side will look like.

Evidence becomes a by-product of working. More and more work now happens inside AI tools and agents, and those tools see the process: the decisions, the versions, what was accepted and what was rejected, how things turned out. A person does not need to publish anything for a record of their judgement to exist. The question becomes who holds that record and who may read it. Ilanooo's proposed making programme is designed around this: it would record the child's decisions, not her keystrokes. The Direction Report already works this way. It goes to the parent, nothing is made public without approval, and a parent can ask for a copy, a correction or deletion. [38]

Institutions change from issuers into witnesses. They will not disappear. A school, an employer or a client will sign specific, checkable statements of what happened (*she shipped this, it is in use; he worked here three years and we would hire him again*) and the person's agent will carry them. This is how credit reporting already works: lenders report what you did to the bureaus, and the judging happens elsewhere. The technology for this, signed credentials kept by the holder, exists today. [34] Its adoption by employers is still thin, and nobody yet has evidence that it changes hiring decisions.

The decisive moment gets checked live. Checking a person directly is now cheap. Brian Jabarian and Luca Henkel ran a field experiment at a recruitment firm in the Philippines with about 70,000 applications for entry-level jobs. Applicants interviewed by an AI voice agent received about 12 per cent more job offers than those interviewed by people, and more of them started and stayed, with human recruiters still making every hiring decision. [41] Universities are moving the same way. The University of Sydney now splits assessment into a secured lane, where students show what they can do without AI under supervision, and an open lane where AI is part of learning. [42]

Degrees shrink but survive. A degree takes years, and time is expensive to fake. So it will keep signalling what it always truly signalled (that a person stayed the course) while losing what it claimed to show, which was skill. The economist Bryan Caplan argued that much of what employers pay for in a degree was always this kind of signal. [43] For people, expect a long mixture: work records, signed statements from witnesses, live checks, and degrees, side by side, for at least a decade.

Democracy and scoring: who computes trust

When people hear "trust infrastructure", many think of China's social credit system and recoil. That reaction is right about the danger and often wrong about the facts.

Jeremy Daum of China Law Translate, who reads the system's documents in the original, wrote in 2018 that despite its name social credit "isn't really focused on scores of any kind": it works mostly through government red lists and blacklists that enforce existing laws, with consequences in particular industries. [44] Genia Kostka's survey across China found surprisingly high approval of these systems, highest among wealthier, urban, older and more educated respondents, with the caution that people may not answer freely under an authoritarian government. [45]

The danger is real all the same, and democracies have named it. In 2014 Danielle Citron and Frank Pasquale described *the scored society*: automated scores that decide people's chances, made by systems no one can inspect or challenge. They argued that a scored person needs to be able to test the score and dispute it. [46] The European Union went further. Article 5 of its AI Act bans social scoring: rating people on their social behaviour or personal traits in a way that leads to unfair treatment in an unrelated part of their life. [40] Credit reporting earned its place only because it was governed, with standard data and rules about accuracy. The scored society is what the same machinery becomes when nobody governs it, and a whole life ends up as one number that follows a person everywhere.

This is where the next era offers a real choice. There are two ways to compute trust with agents.

One authority scores everyone. A government or a platform holds the data, runs the formula and publishes the number. Everyone games the same number. Whoever controls the formula controls who gets in.

Each person's agent works it out for them. The evidence is open or carried by the people it describes. Each chooser's agent reads it and weighs it by what that chooser cares about. There is no single number to game, because there is no single number.

The first is the scored society. The second is closer to how trust worked in the village, where each person made up their own mind from what they saw, now with a reach the village never had. A democracy should want the second, and should write rules that keep the first from creeping in through the back door: through a dominant platform, a single credential that every employer demands, or a default that nobody chose.

Mobility: the opening and the lock-in

I care about trust infrastructure for one reason above the others. It decides who gets a chance.

Every filter we built had to use what it could see. What it could see was what a person's environment had allowed them to show: the right school, the right accent, the polish that comes from never having been told you are average. The filter calibrated itself to that and then called the result intelligence. The child in a low-resource home did not have fewer abilities. She had the same abilities, many of them never expressed, and the filter could not see them.

The next era makes a different filter possible. If checking becomes cheap, a chooser no longer has to rely on the proxy. When a person chooses to show what she actually did, the chooser's agent can read it. A capable person without the right degree or the right network becomes visible on real evidence. That could be one of the largest openings for mobility in a long time.

It could also become a new lock. Three things decide which.

Who owns the work record. If employers and platforms own the record of a person's work, reputation stays locked inside them, the way a driver's ratings cannot be carried from one ride app to another. People become less mobile than today, not more. If the person holds the record and shows it when they choose, mobility opens. Who should hold it, the person's own agent or a neutral third party, is still an open question. I lean to the person's own agent. In *Every Child Will Have an Agent* I asked whose side a child's agent is on: the platform's, the school's, or the child's. [47] The same question decides whether the record of a working life belongs to the person who lived it.

Whether newcomers get a trial. Evidence-based trust rewards long records. A newcomer has none. If agents always choose the option with the most proof, they will lock in a new set of incumbents, and the power law freezes again. The fix is old and it works. Amanda Pallais hired 952 randomly chosen workers on an online platform for a small job, and rated their work publicly. For the inexperienced workers among them, that one job and a short public rating almost tripled their chance of finding any work in the next two months: 30 per cent found work, against 12 per cent of similar workers who were not hired. Their earnings nearly tripled too. [48] A fair trust infrastructure builds in trial channels: a share of every agent's choices spent on promising unknowns, not only on the proven.

Whether trajectory counts, not only position. Meritocracy read a person once, at eighteen, and froze the result. What matters in a world where skills expire quickly is how fast someone is growing and what they keep adding. I have called a society that reads that *growthocracy*: aristocracy was the story of birth, meritocracy the story of ability,

and the next system can be the story of becoming. An agent that reads a person's record over time can see slope, not only height. Whether anyone builds it to do so is a choice.

Seven principles for a trust infrastructure worth having

Pulling the argument together, these are the principles any trust system for the agent era should be held to. Most are not new. What is new is that they are now possible to build.

1. **Prefer evidence that is expensive to fake.** Time, independent agreement, outcomes and live demonstration over anything that can be produced in seconds.
2. **Put the cost back on whoever issues the signal.** A witness, reference or credential means more when its issuer loses something if it is wrong.
3. **Let the person control the record of their work.** Shown by their choice, portable between employers and platforms. (Whether it lives in their own agent or with a neutral holder is open.)
4. **Let many choosers compute, never one authority.** Trust worked out by each person's agent, not a single score that everyone must chase.
5. **Never turn a reading into an entry requirement.** A measure that becomes a condition for admission or hiring gets rehearsed, and families start paying for the rehearsal. [38] That is why Ilanooo's Direction Report is never a credential. A record a person chooses to show is a different thing from a score others demand, and the line between them has to be held on purpose.
6. **Keep room for newcomers.** Trial channels, and weight on trajectory as well as track record.
7. **Make every score inspectable and every error correctable.** The kind of rules credit reporting got half a century ago, and the due process Citron and Pasquale asked for, applied to everything an agent relies on. [7] [46]

What we do not know

Much of this paper is argument, and some of it may be wrong. In the same spirit as Ilanooo's technical paper, here is what the evidence does not yet show.

- **Whether agent checking is accurate and fair.** Checking is getting cheaper. The work cited in chapters 4 and 5 found machine checkers favouring their own writing, swayed by where an item sits on a page, and no better than people at spotting fake reviews. Whether agents can check better than the stamps they replace is not yet shown.
- **How agents choose in the real world.** Most evidence about agent choice comes from simulated markets and preprints. We do not yet have independent field data on what agents buy when real money is at stake, or how much their choices concentrate on incumbents over months.
- **What employers switch to.** The studies in chapter 4 show written signals losing value. None yet measures, at scale, what employers rely on instead (referrals, trials, live interviews, pedigree) or who gains and who loses from the switch.
- **Whether outcome memory can be trusted.** If agents' memories of outcomes replace reviews, someone will try to poison those memories. No one has shown a design that survives that.
- **Whether signed credentials change decisions.** The standards exist and governments are building wallets. There is little evidence yet that employers or schools treat a signed credential differently from a paper one.
- **Whether process records can themselves be faked.** If evidence of judgement comes from the record of a person's work with AI, an agent could be asked to produce a convincing fake record. How expensive that is to do, and to catch, is unknown.
- **The cost to mobility of the transition.** That trust narrows to networks while signals are cheap is plausible and fits the evidence in chapter 5. No study has yet measured its effect on who gets hired, admitted or funded.
- **Education, hiring and markets have been studied apart.** The same mechanism runs through all three, and almost no research looks at them together.

Where this leaves us

Every trust infrastructure humans made was a bargain between scale and truth. The village could not be faked and could not grow. The stamp could grow and measured the wrong thing. The rating could grow and could be gamed. AI broke the stamp and the rating by making them free to produce. It also gave us, for the first time, a way to check at the scale of the stamp with something close to the truth of the village.

Whether we use it that way is not decided by the technology. It is decided by who holds the record, who computes the trust, and whether the newcomer gets a chance. The card networks are deciding how an agent proves who sent it. The European Commission is deciding what goes into a citizen's wallet. Hiring software is deciding what an interview records. Whoever builds children's agents is deciding whose side those agents are on.

At Ilanooo we work on the part that comes before all of it: people who can direct the machine rather than be carried by it, and who leave a record of real things made for real people. A child who grows up making things others use builds a record she can choose to show. A child who only consumes leaves none.

Trust was always made of checking. Now that checking is cheap, the question is no longer whether we can trust strangers at scale. It is whether the new way of doing it will see the people the old one could not.

References

1. Firoz Azees, "The Collapse of Talent Signals," ilanooo.com/direction/talent-signal-collapse, 10 September 2026.
2. Robert J. Sampson, Stephen W. Raudenbush and Felton Earls, "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy," *Science* 277(5328), 1997.
3. R. I. M. Dunbar, "Neocortex size as a constraint on group size in primates," *Journal of Human Evolution* 22(6): 469–493, 1992.
4. Avner Greif, "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *American Economic Review* 83(3): 525–548, 1993.
5. Jeremy Edwards and Sheilagh Ogilvie, "Contract enforcement, institutions, and social capital: the Maghribi traders reappraised," *Economic History Review* 65(2): 421–444, 2012.
6. Lynne G. Zucker, "Production of trust: Institutional sources of economic structure, 1840–1920," *Research in Organizational Behavior* 8, 1986.
7. FICO, "About us" (founded 1956; first FICO score 1989), fico.com; and the US Fair Credit Reporting Act, enacted 26 October 1970.
8. Michael Young, *The Rise of the Meritocracy*, Thames & Hudson, 1958.
9. Lauren A. Rivera, *Pedigree: How Elite Students Get Elite Jobs*, Princeton University Press, 2015.
10. Rachel Botsman, *Who Can You Trust? How Technology Brought Us Together and Why It Might Drive Us Apart*, PublicAffairs, 2017.
11. Nan Hu, Jie Zhang and Paul A. Pavlou, "Overcoming the J-shaped distribution of product reviews," *Communications of the ACM* 52(10): 144–147, 2009.
12. Apostolos Filippas, John J. Horton and Joseph M. Golden, "Reputation Inflation," *Marketing Science* 41(4): 733–745, 2022.
13. Michael Anderson and Jeremy Magruder, "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *The Economic Journal* 122(563): 957–989, 2012.
14. Michael Spence, "Job Market Signaling," *Quarterly Journal of Economics* 87(3): 355–374, 1973.
15. Amotz Zahavi, "Mate selection: A selection for a handicap," *Journal of Theoretical Biology* 53(1): 205–214, 1975.
16. Alan Grafen, "Biological signals as handicaps," *Journal of Theoretical Biology* 144(4): 517–546, 1990.
17. George A. Akerlof, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84(3): 488–500, 1970.
18. Anaïs Galdin and Jesse Silbert, "Making Talk Cheap: Generative AI and Labor Market Signaling," [arXiv:2511.08785](https://arxiv.org/abs/2511.08785), 2025 (working paper).
19. Jingyi Cui, Gabriel Dias and Justin Ye, "Signaling in the Age of AI: Evidence from Cover Letters," [arXiv:2509.25054](https://arxiv.org/abs/2509.25054), 2025 (working paper).
20. Peter Scarfe, Kelly Watcham, Alasdair Clarke and Etienne Roesch, "A real-world test of artificial intelligence infiltration of a university examinations system: A 'Turing Test' case study," *PLOS ONE* 19(6): e0305354, 2024.
21. Weiyao Meng and colleagues, "Large Language Models as 'Hidden Persuaders': Fake Product Reviews are Indistinguishable to Humans and Machines," [arXiv:2506.13313](https://arxiv.org/abs/2506.13313), 2025 (preprint; small study).
22. Jiannan Xu, Gujie Li and Jane Yi Jiang, "AI Self-preferencing in Algorithmic Hiring: Empirical Evidence and Insights," [arXiv:2509.00462](https://arxiv.org/abs/2509.00462), 2025 (preprint).
23. Matt Sigelman, Joseph Fuller and Alex Martin, *Skills-Based Hiring: The Long Road from Pronouncements to Practice*, Burning Glass Institute with the Harvard Business School Managing the Future of Work project, February 2024.

24. Amanda Pallais and Emily Glassberg Sands, "Why the Referential Treatment? Evidence from Field Experiments on Referrals," *Journal of Political Economy* 124(6): 1793–1828, 2016.
25. Edelman Trust Institute, *2026 Edelman Trust Barometer: Trust Amid Insularity*, 2026.
26. Gallup, "U.S. Public Trust in Higher Ed Rises From Recent Low," Lumina Foundation–Gallup survey, 15 July 2025.
27. Amine Allouah, Omar Besbes, Josué D. Figueroa, Yash Kanoria and Akshit Kumar, "What Is Your AI Agent Buying? Evaluation, Biases, Model Dependence, & Emerging Implications for Agentic E-Commerce," arXiv:2508.02630, 2025 (preprint).
28. Xi Chu and Yupeng Hou, "Incumbent Advantage: Brand Bias and Cognitive Manipulation Dynamics in LLM Recommendation Systems," arXiv:2606.17443, 2026 (preprint).
29. Mahammed Kamruzzaman, Hieu Minh Nguyen and Gene Louis Kim, "'Global is Good, Local is Bad?': Understanding Brand Bias in LLMs," *Proceedings of EMNLP 2024*: 12695–12702.
30. Firoz Azees, "Agents Remember: How the Power Law Freezes Again," ilanooo.com/essays/agents-remember-how-the-power-law-freezes-again, 16 October 2026.
31. Linux Foundation, "Linux Foundation Launches the Agent2Agent Protocol Project to Enable Secure, Intelligent Communication Between AI Agents," 23 June 2025.
32. Visa, "Visa Introduces Trusted Agent Protocol: An Ecosystem-Led Framework for AI Commerce," 14 October 2025; Mastercard, "Mastercard unveils Agent Pay," 29 April 2025.
33. Alan Chan and colleagues, "Visibility into AI Agents," *ACM FAccT 2024* (arXiv:2401.13138); Alan Chan and colleagues, "Infrastructure for AI Agents," *Transactions on Machine Learning Research*, 2025 (arXiv:2501.10114).
34. W3C, *Verifiable Credentials Data Model v2.0*, W3C Recommendation, 15 May 2025.
35. Regulation (EU) 2024/1183 (European Digital Identity Framework) and the European Commission's EU Digital Identity Wallet pages.
36. Ministry of Electronics and Information Technology, DigiLocker (digilocker.gov.in); Ministry of Education, APAAR (apaar.education.gov.in).
37. Coalition for Content Provenance and Authenticity, *C2PA Technical Specification*, version 2.4, 2026.
38. Firoz Azees, *Teaching Children to Direct AI: A Method, a Record and a Trial*, ilanooo, ilanooo.com/teaching-children-to-direct-ai, October 2026.
39. Jakob Nielsen, "The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities," Nielsen Norman Group, 8 October 2006.
40. Regulation (EU) 2024/1689 (the AI Act), Article 5(1)(c) and Annex III, points 3 and 4.
41. Brian Jabarian and Luca Henkel, "Voice AI in Firms: A Natural Field Experiment on Automated Job Interviews," SSRN 5395709 (2025) and arXiv:2607.28222 (2026), working paper.
42. Danny Liu and Adam Bridgeman, "Embracing the future of assessment at the University of Sydney," Teaching@Sydney, December 2023.
43. Bryan Caplan, *The Case against Education: Why the Education System Is a Waste of Time and Money*, Princeton University Press, 2018.
44. Jeremy Daum, "The redlists are coming! The blacklists are coming!," China Law Translate, 30 March 2018.
45. Genia Kostka, "China's social credit systems and public opinion: Explaining high levels of approval," *New Media & Society* 21(7): 1565–1593, 2019.
46. Danielle Keats Citron and Frank Pasquale, "The Scored Society: Due Process for Automated Predictions," *Washington Law Review* 89: 1, 2014.
47. Ilanooo, "Every Child Will Have an Agent. Whose Side Is It On?," ilanooo.com/learning/every-child-will-have-an-agent-whose-side-is-it-on, 16 October 2026.
48. Amanda Pallais, "Inefficient Hiring in Entry-Level Labor Markets," *American Economic Review* 104(11): 3565–3599, 2014. The 12 and 30 per cent figures are from the NBER Working Paper 18917 version (2013), section on the coarse evaluation treatment.