When Is a Voice an Infrastructure?

Care, Classification, and the Politics of Who Answers

Johan Michalove*

February 26, 2026 — *Working draft*

The world is always breaking, carrying much of what we care about into loss and oblivion. There is nothing and no one to save us. We are always and everywhere alone, but for the profuse and teeming worlds around us.

Now let us get to work.

— Steven J. Jackson, "Ordinary Hope"

Abstract

When someone in New York City calls a phone number and asks for help finding shelter, food, or medical care, what answers? For over a century, the answer was a human: a switchboard operator, a social worker, a 211 specialist, a volunteer on a warm line. Increasingly, the answer is an algorithm. This paper examines the emergence of voice-based artificial intelligence in social services through three analytical lenses drawn from science and technology studies: *infrastructure*, *care*, and *testing*. Drawing on Susan Leigh Star and Karen Ruhleder's relational definition of infrastructure, I ask not *what* a voice AI system is but *when* it becomes an infrastructure of care—and for whom. Through an analysis of the 211 Human Services Indexing System (10,500+ terms classifying the full range of human need), I apply Geoffrey Bowker and Susan Leigh Star's framework of classification to show how taxonomies of need shape what an AI can and cannot recognize. Finally, extending Trevor Pinch's sociology of testing, I argue that every call to a voice AI helpline is simultaneously a test of the system and an act of care-seeking, and that these two functions exist in productive tension. The paper contributes to ongoing conversations in STS about automation and inequality, the politics of infrastructure, and the ethics of care in technoscience, while offering a historically grounded analysis of the telephone as a technology whose deepest social function has always been care.

*Department of Information Science, Cornell University. jam844@cornell.edu. For Trevor Pinch (1952–2022), who taught that testing is never just technical.

Keywords: infrastructure, care, voice AI, classification, telephone, social services, testing

1 Introduction: Number, Please

A phone rings. Someone in New York City needs help. They are elderly and their heat has been shut off, or they are newly arrived and do not know where to find food, or they are leaving a dangerous situation and need somewhere safe to sleep tonight. They pick up a telephone—in 2026, still the most reliable communication technology available to people in crisis—and dial a number. What answers?

For the better part of a century and a half, the answer was a human voice. The earliest telephone operators greeted callers with "Number, please?"—a phrase so minimal it barely constituted a greeting, yet so effective it became the template for an entire industry of voice-mediated care (Fischer, 1992). From Emma Nutt, the first female telephone operator hired in Boston in 1878, through the professionalization of social work in the Progressive Era, through the grassroots information and referral movement of the 1960s and the creation of 211 in 2000, the telephone has been a technology of connection, triage, and care. When you called, a person answered. That person listened, asked what you needed, and tried to help.

This is changing. Voice-based artificial intelligence systems are entering the space historically occupied by human operators, crisis counselors, and social service specialists. The question this paper asks is not whether such systems should exist—they already do—but what it means, in the vocabulary of science and technology studies, for a voice to become an infrastructure of care. I borrow the framing from Star and Ruhleder (1996), whose relational definition of infrastructure insists that we ask not *what* an infrastructure is but *when* something becomes infrastructure in relation to organized practices. A telephone line is not inherently care infrastructure. It becomes so when a woman uses it to check on her elderly mother (Fischer, 1992), when a 211 specialist connects a caller to a warming center, when a voice AI system searches a database of 835 community resources and reads the results aloud to someone who cannot use a smartphone. It ceases to be infrastructure—and becomes an obstacle—when the caller's need falls outside the taxonomy, when the referral is broken, when the synthetic voice cannot understand the caller's accent.

The paper develops three analytical lenses. The first is **infrastructure**: drawing on Star and Ruhleder, Larkin (2013), Jackson (2014), and Mattern (2018), I examine how voice AI systems become (and fail to become) care infrastructure, attending to both the politics and the poetics of the synthetic voice. The second is **classification**: applying Bowker and Star (1999) to the 211 Human Services Indexing System—a taxonomy of over 10,500 terms that classifies the full range of human need into ten categories across six hierarchical tiers—I

analyze how the categories through which an AI recognizes need are themselves political instruments that valorize some forms of suffering and silence others. The third is **testing**: extending Trevor Pinch's sociology of testing (Pinch, 1993), I argue that every real call to a voice AI helpline is simultaneously a current test (does this system work?) and a prospective test (should this system exist?), and that the populations whose needs are most urgent are also the populations on whom the system is, effectively, being tested.

These lenses converge on a single question: *who gets a human, and who gets a machine?* This is not a technical question about system design but a political question about the allocation of care (Tronto, 2013). When a city deploys voice AI to handle social service calls, it is making a decision—often implicit, often invisible—about which populations merit the attention of a human being and which can be routed to an algorithm. The paper traces the historical, theoretical, and ethical dimensions of this decision.

A note on method and disclosure. This paper is written from the position of a researcher who has designed and deployed a voice AI system for community resource navigation in New York City. The analysis draws on the design methodology—its grounding in evidence-based practice across social work, crisis intervention, and information and referral standards—rather than on the system's technical architecture, which is not disclosed here. The contribution is analytical, not prescriptive: I do not argue that voice AI *should* replace human operators, nor that it *should not*. I argue that the phenomenon demands the kind of attention that STS is equipped to provide—attention to the relational, the classificatory, and the political dimensions of what it means for a voice to answer when someone calls for help.

The paper proceeds as follows. Section 2 traces the history of the telephone as a care technology, from the feminized labor of switchboard operators through the creation of 211 and the emergence of warm lines. Section 3 develops the theoretical framework of voice-as-infrastructure, drawing on infrastructure studies, repair, and maintenance. Section 4 analyzes the politics of classification in social service taxonomies. Section 5 examines the automation of care labor and its implications for inequality. Section 6 applies Pinch's sociology of testing to the deployed voice AI system. Section 7 draws on Jackson (2023) to consider what it means to build with ordinary hope in broken worlds. Section 8 concludes.

2 The Telephone as Care Technology

The telephone was not designed to be a care technology. Alexander Graham Bell imagined it as a business instrument. The Bell System marketed it for commerce, emergencies,

and household management. But from the earliest years of its adoption, Americans—and especially American women—used the telephone for something else entirely: *sociability* (Fischer, 1992). They called to check on elderly relatives, to coordinate childcare, to organize church suppers, to share news of illness and death, to alleviate the isolation of rural and suburban domestic life. Fischer's central finding is that the telephone reinforced and deepened social patterns that already existed rather than creating new ones. It was, in his phrase, a technology that people bent to their own purposes—and the purpose they chose, against the wishes of the industry, was care.

This history matters because it establishes a principle that voice AI designers ignore at their peril: the social meaning of a telephone technology is not determined by its designers but negotiated by its users. Pinch and Bijker (1984) would recognize this as a case of *interpretive flexibility*—the same artifact means different things to different relevant social groups. The Bell System saw a business tool; women in rural Illinois saw a lifeline. The users won. By the 1920s, the industry had abandoned its campaign against "frivolous" social calls and begun marketing the telephone as a technology of domestic connection (Fischer, 1992).

The labor that made this connection possible was gendered from the start. The first telephone operators were teenage boys, recruited from telegraph offices; they proved "impatient and rude, full of pranks, including disconnecting customers and misdirecting their calls." Emma Nutt, hired in Boston in 1878, was the first female operator—"patient and savvy, with a voice that was cultured and soothing." Within a decade, telephone operating had become an exclusively female profession. Green (2001) traces how this gendering was also racialized: the Bell System's hiring practices systematically excluded Black women from operator positions until the 1960s, even as the labor of connection—the emotional work of making callers feel heard and helped—was coded as naturally feminine. Hochschild (1983) would later theorize this as *emotional labor*: the management of feeling as part of waged work, the production of a particular affective state (warmth, patience, attentiveness) as a condition of employment.

The operator's greeting—"Number, please?"—was a masterpiece of compressed care. It announced the operator's role, invited the caller to state their need, and offered immediate attention, all in two words. The Bell System's "voice with a smile" training philosophy, developed under Theodore Vail's leadership, taught operators that voice quality and emotional tone were inseparable from service quality. The voice *was* the service. This insight—that in a voice-mediated system, the affective qualities of the voice are not incidental to the service but constitutive of it—remains the most important design principle for any voice AI system, and it cannot be found in any machine learning paper.

2.1 From operators to information and referral

The professionalization of telephone-based social services followed a parallel track. In the 1960s and 1970s, community-based information and referral (I&R) services emerged across the United States as a grassroots response to the fragmentation of social services. These services—staffed by trained specialists, often housed in United Way agencies or community action programs—provided a single point of contact for people navigating complex webs of housing, food, healthcare, legal aid, and other needs. The Alliance of Information and Referral Systems (AIRS, now Inform USA) was established in the early 1970s to create national standards.

In June 1997, the United Way of Greater Atlanta launched the nation's first 211 contact center—a 24/7, multilingual referral service accessible by dialing three digits. In July 2000, the Federal Communications Commission approved 211 for nationwide use (Federal Communications Commission, 2000). Today, approximately 95% of the U.S. population has access to 211 services through over 200 agencies. The canonical I&R process, codified in AIRS standards, follows five stages: (1) opening and rapport, (2) assessment through active listening, (3) clarification and mutual understanding, (4) referral with enough information for informed choice, and (5) closing with follow-up offer.

The 211 system represents the institutionalization of the telephone as care infrastructure—the formalization of what Fischer's rural callers had been doing informally for decades. But institutionalization brought classification, and classification brought its own politics (Section 4).

2.2 Warm lines and the relational turn

A third tradition of telephone-based care emerged from the consumer/survivor movement in mental health during the 1980s and 1990s. *Warm lines*—distinct from crisis *hotlines*—are typically staffed by peers with lived experience of mental health challenges, and serve callers who are not in immediate danger but need connection, emotional support, or help navigating resources. Research has found that warm line users report decreased use of crisis services, reduced psychiatric hospitalizations, and increased community integration.

The warm line innovation was not technological but relational: the shift from professional clinical response to peer support. The most important thing about a warm line is not what it does (provide information) but *who answers* (someone who has been there). This poses a fundamental challenge to voice AI: an algorithm has no lived experience. It cannot be a peer. What, then, does it offer? The answer, I will argue, lies not in simulating

peer experience but in maintaining the infrastructure through which care can flow—an infrastructure that is always breaking, always requiring repair (Jackson, 2014), and always in need of the kind of ordinary, patient attention that Jackson (2023) calls hope.

3 When Is a Voice an Infrastructure?

Star and Ruhleder (1996) propose a relational definition of infrastructure that has become foundational in STS: "infrastructure appears only as a relational property, not as a thing stripped of use." Their central question—"*when*—not *what*—is an infrastructure?"—insists that something becomes infrastructure only in relation to organized practices. A railroad is infrastructure for a commuter; it is a barrier to the community whose neighborhood it bisects. A database of 835 community resources is infrastructure for the caller who is connected to a warming center; it is an obstacle for the caller whose need does not appear in any category.

This relational definition is essential for analyzing voice AI in social services because it refuses the assumption—common in both technology journalism and policy discourse—that a voice AI system simply *is* infrastructure by virtue of being deployed. The question is not whether the system exists but for whom, under what conditions, it functions as infrastructure. Star (1999) identifies several dimensions of infrastructure that illuminate this distinction: infrastructure is *embedded* in other structures; it is *transparent to use* (invisible when it works, visible when it breaks); it has *reach or scope* beyond a single event; it is *learned as part of membership* in a community of practice; and it is *built on an installed base*, inheriting the strengths and limitations of what came before.

Voice AI in social services inherits the installed base of 211—its taxonomy, its standards, its conception of what constitutes a "human service." It is embedded in the infrastructure of the telephone network, which is itself embedded in the infrastructure of the city. It is transparent to use when a caller receives a useful referral without thinking about how the system works; it becomes visible—painfully, sometimes devastatingly visible—when the system fails to understand the caller's speech, provides a broken referral, or terminates the call prematurely.

3.1 The politics and poetics of voice

Larkin (2013) argues that infrastructure has both *political* dimensions—how it distributes power, resources, and access—and *poetic* dimensions—how it generates affect, desire, and aesthetic experience. He challenges the common STS claim that infrastructure is invisible

until it breaks down, arguing that this is "flatly untenable": many infrastructures are deliberately spectacular, symbolic, or affective. The voice that answers the phone when someone calls for help is not merely a functional interface. It is an aesthetic and affective experience—a poetic dimension of infrastructure that is as politically significant as its routing logic.

What does it mean to call for help and hear a synthetic voice? The quality of that voice—its warmth or coldness, its naturalness or uncanniness, its pace and patience—communicates something about who is being served and how much their care matters. The Bell System understood this in the 1880s: the voice *is* the service. UNESCO (2019) has documented how the default feminization of AI voice assistants—Siri, Alexa, Cortana—reproduces the gendered servility that Hochschild (1983) identified in human emotional labor. The choice of a voice model for a social service system is not a technical decision about audio quality; it is a decision about what register of care the system communicates.

Larkin's framework also illuminates the temporal politics of voice infrastructure. When a caller is placed on hold—listening to music while an algorithm searches a database—they are experiencing the system's temporality imposed on their own. Virilio (2006) would call this a form of dromological power: the politics of speed and waiting, of whose time is valued and whose is not. Wajcman (2015) and Crary (2013) have analyzed how digital capitalism restructures temporal experience; the voice AI helpline participates in this restructuring by converting the open-ended temporality of a human conversation into the algorithmic temporality of search, synthesis, and response. The caller waits. The algorithm does not experience waiting. The asymmetry is not incidental to the system; it is constitutive of it.

3.2 Repair and maintenance

Jackson (2014) proposes "broken world thinking" as an alternative to the productivist bias of technology studies: "Breakdown, dissolution, and change, rather than innovation, development, or design as conventionally practiced and thought about, are the key themes and problems facing new media and technology scholarship today." He argues that maintenance and repair are unheralded sites of creativity, knowledge, and "a neglected ethics of care."

A voice AI system for social services is, fundamentally, a repair project. The resource database changes constantly: organizations close, phone numbers change, eligibility criteria shift, hours of operation fluctuate with seasons and funding cycles. A warming center opens when the temperature drops below 32 degrees and closes when it rises above.

A food pantry exhausts its supply and cannot serve callers until next week. A shelter reaches capacity at 9 PM and turns people away. The system's data is always partially broken, and the work of maintaining it—verifying that the food pantry is still open, that the phone number still works, that the address is correct—is itself a form of care labor.

Mattern (2018) extends this argument to data specifically, arguing that the work of curating, cleaning, and maintaining data is a form of care labor that is "invisible, gendered, undervalued, and essential." The 211 operators, librarians, and data archivists who maintain the resource databases on which voice AI systems depend are *maintainers* in Mattern's sense: their labor is constitutive of the system's capacity to function, but it is systematically excluded from accounts of the system's "intelligence." The AI gets the credit; the maintainer gets the wage.

This is why Jackson's rethinking of repair is not merely a theoretical contribution but an ethical one. If the real work of care infrastructure is not designing the system but maintaining it, then the question of who maintains—who checks the data, who verifies the referrals, who updates the taxonomy—is as politically consequential as the question of who designs. Broken referrals are not bugs to be fixed; they are the system's primary condition. Building voice AI for social services means committing to repair as a permanent practice—a commitment that is, in Jackson's vocabulary, an act of care.

4 Matters of Care, Matters of Classification

4.1 The four phases of care

Tronto (1993) defines care as "a species activity that includes everything that we do to maintain, continue, and repair our 'world' so that we can live in it as well as possible." She identifies four analytically distinct phases: (1) *caring about*—recognizing that care is necessary; (2) *taking care of*—assuming responsibility for meeting the need; (3) *care-giving*—the actual work of care; and (4) *care-receiving*—the response of the person being cared for. Tronto insists that care is not a sentimental feeling but a political practice, and that the boundaries between ethics and politics, public and private, are precisely the boundaries that keep care invisible and devalued.

When care is delegated to a machine, which of these phases are preserved and which collapse? A voice AI system can perform aspects of care-giving—searching a database, formulating a response, providing referrals. But can it perform *caring about*—the recognition that this particular person, in this particular situation, needs something that cannot be reduced to a database query? And who is responsible when the *care-receiving* phase

fails—when the referral is broken, the shelter is full, the resource no longer exists?

Puig de la Bellacasa (2011) extends Latour's concept of "matters of concern" into a more demanding framework: "matters of care." Where Latour moves from "matters of fact" to "matters of concern," Puig de la Bellacasa argues that concern is not equivalent to care. Care requires not just attention but "a speculative commitment to neglected things"—an active orientation toward generating and sustaining livable worlds. The point is not only to expose invisible labors of care but to *generate* care. The distinction is crucial for this analysis: a voice AI system for social services is not merely a "matter of concern" (something to worry about, study, and debate) but a "matter of care" (something that requires active commitment to making it work for the populations it serves).

The Care Collective (2020) extend this argument to the political domain, arguing for care as a form of infrastructure—not metaphorically but materially. Care requires institutions, resources, and sustained attention. When these are withdrawn—through austerity, through privatization, through the quiet replacement of human workers with algorithms—the infrastructure of care erodes, and the populations most dependent on it are the first to suffer.

4.2 Sorting things out

Bowker and Star (1999) argue that classification systems are not neutral descriptions of the world but moral and political instruments: "each standard and category valorizes some point of view and silences another." Through case studies of the International Classification of Diseases, the Nursing Interventions Classification, and race classification under South African apartheid, they show that classification systems produce advantage or suffering, create and destroy jobs, and make the political choices embedded in categories disappear from view through the process of naturalization.

The 211 Human Services Indexing System (211HSIS) is a controlled vocabulary containing over 10,500 terms organized in a hierarchy of up to six tiers across ten basic service categories: Basic Needs; Consumer Services; Criminal Justice and Legal Services; Education; Environment and Public Health/Safety; Health Care; Income Support and Employment; Individual and Family Life; Mental Health and Substance Abuse Services; and Organizational/Community/International Services. This taxonomy determines what counts as a "human service," what counts as a "need," and—crucially—what falls outside the classification entirely.

When a voice AI system receives a call, it must perform a classificatory act: it must map the caller's spoken description of their situation onto the categories available in

the resource database. This mapping is not transparent. A caller who says "I need food" might be experiencing income loss, benefits lapse, domestic violence, disability, immigration-related barriers, or any combination of these. The presenting problem, as social work practice has long recognized, is frequently not the underlying need. "I need food" may mean "I just left my partner and I have nothing." The 211HSIS taxonomy, with its 10,500 terms, cannot capture this; it can only classify the surface.

Douglas (1966) provides an even more fundamental analysis of classification: categories produce not only inclusion but *exclusion*, and what falls outside the classificatory system is rendered anomalous, dangerous, polluting. Scott (1998) extends this to the state, arguing that the drive to make populations *legible*—to render complex social realities into simplified administrative categories—is a constitutive feature of modern governance. The 211 taxonomy is a legibility project: it makes human suffering administratively navigable. But what it cannot classify, it cannot see.

For voice AI, the classificatory problem is compounded by the translation from spoken language to database category. Human 211 specialists—trained in active listening, motivational interviewing, and the recognition of underlying needs—can hear "I need food" and ask the follow-up question that opens the door to the real situation. An AI, operating within the constraints of its training and system instructions, performs a different kind of classification: one that is faster, more consistent, and less able to hear what is not said.

5 The Algorithmic Operator

Eubanks (2018) introduces the concept of the "digital poorhouse"—"an invisible institution that profiles, polices, and punishes the poor when they come into contact with public services." Through three case studies—Indiana's automated welfare eligibility system, the LA coordinated entry system for homeless services, and Allegheny County's Family Screening Tool—she demonstrates that automated systems do not neutrally deliver services but encode institutional biases, treating vulnerable populations as data problems to be managed rather than as people with rights.

The Indiana case is paradigmatic. After Governor Mitch Daniels signed a $1.3 billion contract with IBM to automate welfare eligibility determination, the system denied over one million applications in three years, interpreting any procedural error—a missed appointment, a document filed one day late—as "failure to cooperate." Omega Young of Evansville, Indiana, had her Medicaid benefits terminated via automated determination while she was hospitalized with terminal cancer. She won her appeal the day after she died.

Eubanks's work establishes a critical distinction that any analysis of voice AI in social services must confront. There is a difference between a system that *determines eligibility*—that decides who gets help—and a system that *routes callers to resources*—that helps people find help. The former is gatekeeping; the latter is wayfinding. A voice AI system that searches a database and reads results aloud is, architecturally, closer to a search engine than to an eligibility determination system. But the question is whether this distinction holds under pressure. When the resource database is incomplete, when certain communities' needs are underrepresented, when the AI's speech recognition fails for some accents more than others, the routing system begins to function as a gatekeeping system—not by design, but by default.

Benjamin (2019) names this mechanism: "default discrimination"—discrimination that grows not from malicious intent but from "socially and historically ignorant design processes." She defines the "New Jim Code" as "the employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era." Noble (2018) demonstrates how this operates in search engines specifically, coining the term "technological redlining" for new modes of racial profiling enacted through algorithms.

A voice AI social service system is, architecturally, a search engine. Someone calls, describes a need, and the system searches a database and returns results. Noble's work demands that we ask: whose needs does this search engine serve well, and whose does it serve poorly? The resource database reflects what has been cataloged—which organizations exist, which have been entered into the database, which categories of need have been recognized by the taxonomy. If certain communities' needs are underrepresented because those organizations are less visible, less funded, or less legible to the taxonomy, the AI will reproduce and amplify that gap.

Koenecke et al. (2020) provide direct empirical evidence: commercial automated speech recognition systems from Google, Amazon, Apple, IBM, and Microsoft produce error rates roughly twice as high for Black speakers as for white speakers (35% versus 19%). If a voice AI social service system uses similar speech recognition technology, it will systematically mishear the populations most in need of its services. The technical term is "word error rate"; the social term is exclusion.

Tracey and Garcia (2024) offer the closest existing scholarship to the phenomenon this paper analyzes. Through a qualitative study of 15 care workers in homeless services, they examine how the VI-SPDAT (Vulnerability Index—Service Prioritization Decision Assistant Tool) transforms care labor into data work. Care workers develop *workarounds* to resist the algorithm's reductions—"rephrasing questions or asking prompting follow-up

questions that might raise scores," often at risk of professional sanction. The AI system faces an analogous tension: it must classify the caller's need (to search the database) while also attending to the caller's situation (to provide useful, appropriate help). When the need does not map cleanly to a database category, the AI must improvise. But the AI's "workarounds" are encoded in system instructions rather than emerging from professional judgment, empathy, and lived experience.

Woolgar (1990) provides the STS vocabulary for this problem: "configuring the user." Designers do not simply build systems for pre-existing users; they construct the user through the design process, defining what kinds of interactions are possible and what kinds of requests are legible. A voice AI system configures the caller—it determines what can be asked, in what language, in what register, and what kinds of responses the system can produce. Foucault (1977) would recognize this as a form of disciplinary power: the system produces docile subjects who learn to ask questions the system can answer. Browne (2015) would add that this disciplinary function operates differently on different bodies—that the history of surveillance is a history of the surveillance of Blackness, and that being rendered visible to an automated system carries costs that are unevenly distributed along racial lines.

6 Testing Care

Trevor Pinch's "Testing—One, Two, Three ... Testing!" (1993) argues that testing is not merely a technical procedure for evaluating whether a technology works but a social process through which the meaning, performance, and acceptability of technologies are negotiated. What counts as a "test," who gets to test, what counts as "passing," and whose criteria of success apply—these are social questions, not technical ones. Pinch shows that testing is a site where the interpretive flexibility of a technology—its openness to radically different meanings for different social groups (Pinch and Bijker, 1984)—is temporarily resolved (or not).

Applied to voice AI in social services, Pinch's framework reveals something that neither the technical literature nor the policy discourse adequately acknowledges: every real call is a test. Not a controlled experiment with informed consent and IRB approval, but a test in Pinch's sociological sense—an encounter in which the system's meaning, performance, and acceptability are negotiated between the caller and the technology. The caller does not experience themselves as a test subject; they experience themselves as someone who needs help. But the system learns from the interaction—its designers adjust parameters, rewrite instructions, modify the database in response to what calls reveal—and in

this sense, the caller's need is simultaneously the occasion for care and the material for testing.

This dual function—care and test—creates a tension that Pinch's framework helps articulate. In his earlier work on the sociology of scientific testing, Pinch distinguishes between "current" tests (does this particular instance work as intended?) and "prospective" tests (should this kind of thing exist at all?). Every call to a voice AI helpline is both. The current test asks: did the caller receive a useful referral? Did the system understand the speech? Was the response appropriate? The prospective test asks: should an algorithm be answering this call at all? Is this the right way to provide care? Should society invest in voice AI rather than in human social workers?

Downer (2011) extends the sociology of testing into a sociology of failure, arguing that certain classes of system failure are not merely possible but inevitable—that tightly coupled, complex systems produce "normal accidents" (Perrow, 1984) that no amount of testing can prevent. For voice AI in social services, the relevant failures are not catastrophic (the system is not an aircraft or a nuclear reactor) but cumulative: the referral that leads nowhere, the resource that has closed, the caller who hangs up in frustration, the accent the speech recognition cannot parse. Each individual failure is minor. In aggregate, they constitute a pattern—and the pattern is not random. It falls disproportionately on the populations whose needs are most urgent and whose voices are least well represented in the training data (Koenecke et al., 2020).

The relevant social groups for this technology—in Pinch and Bijker's (1984) SCOT framework—include callers (elderly people, non-English speakers, people in crisis, people experiencing homelessness), social workers, 211 specialists, the system's designer, the AI company, the telephony provider, the organizations listed in the database, city agencies, and funders. Each group interprets the technology differently. For the caller, it may be a lifeline or an indignity. For the city, it may be an efficiency tool or a liability. For the designer, it may be care infrastructure or a compromise. For the AI company, it may be a case study or a revenue source. SCOT insists that none of these interpretations is the "correct" one—the technology is what these negotiations make it. The question is which groups have the power to define success, and which bear the consequences of failure.

7 Ordinary Hope, Ordinary Work

Jackson (2023) offers a speculative argument for the centrality of hope and its intimate connection to the projects of ecological care and repair. He asserts five propositions: that hope is not predictive; that its measure is not accuracy but efficacy; that it may be ex-

pressed in orientations toward change but also in "forms of modest patience and enduring"; that it is above all a property of ordinary work; and that it is a collective accomplishment. Against both the "vertical" hope of messianic traditions (which projects upward toward transcendence) and the critical tradition's hermeneutics of suspicion (which tends to frame hope as false consciousness), Jackson argues for a "horizontal" hope grounded in pragmatist philosophy and the work of repair: "a feet-in-the-mud, dirt-under-fingernails hope, and not one expressed in the plaintive or beseeching gaze towards heaven."

This account of hope—as ordinary, patient, horizontal, and enacted through work rather than expressed through aspiration—provides the ethical orientation this paper has been building toward. Building and maintaining a voice AI system for community resource navigation is an act of ordinary hope. It is not a revolutionary intervention; it does not transform the structural conditions that produce homelessness, hunger, or isolation. It does not replace the human social worker, the community health worker, or the peer volunteer on the warm line. What it does—when it works—is keep a line open. It answers the phone. It searches the database. It reads the results aloud. It says, in a synthetic but not unkind voice, "Here is what I found." And then it asks: "Is there anything else?"

This is modest. It is also, in the vocabulary of care ethics, real: a concrete act of taking-care-of (Tronto, 1993), a speculative commitment to neglected things (Puig de la Bellacasa, 2011), a form of maintenance that is constitutive rather than ancillary (Mattern, 2018). The system is always partially broken. The referrals are always partially wrong. The database is always partially stale. The voice is synthetic, the waiting is algorithmic, and the care is mediated by infrastructure that is relational, political, poetic, and fragile. But the phone rings, and something answers, and for the caller who is connected to a warming center on a night when the temperature is 15 degrees, this is not nothing. It is not a human. But it is not nothing.

Jackson writes: "If one cares about broken worlds, and not in a dystopian or apocalyptic way … one is drawn—inevitably, I believe—to the question of hope." The broken world in question here is the American social safety net—a system in which the presenting problem is frequently not the underlying need, in which only 36% of 211 referrals result in the caller actually receiving assistance (Boyum et al., 2016), in which housing referral success rates hover around 17%, and in which the populations most in need are also the most surveilled (Browne, 2015), the most misheard (Koenecke et al., 2020), and the most likely to be denied care by the very systems designed to provide it (Eubanks, 2018). To build within this broken world—to write code that searches a database and reads the results aloud to someone who called because they need help—is to practice what Jackson calls "thick hope": hope that is "connected to the living of lives and to

patient entanglements with the things, people and places around us."

Shotwell (2016) offers a related framework: *against purity*—living ethically in compromised times. Building AI for social services is inherently compromised. The system uses speech recognition that mishears Black speakers at twice the rate of white speakers. It operates within a taxonomy that cannot capture the full complexity of human need. It routes callers to resources that may no longer exist. It replaces—or supplements, or extends, depending on which relevant social group you ask—the labor of human care workers. There is no pure position from which to build. But Shotwell argues that the demand for purity—the insistence that one must wait until conditions are perfect before acting—is itself a form of privilege. The caller who needs a warming center tonight cannot wait for a structurally just society. The question is not whether the system is perfect but whether it sustains, in Jackson's terms, "more meaningful forms of action and relationality in the world."

This is not an argument for complacency. It is an argument for building while knowing—for building *because* one knows—that the world is broken, the system is imperfect, and the work of repair never ends. "Now let us get to work" (Jackson, 2023).

8 Coda: For Trevor

Trevor Pinch taught a course called Inside Technology at Cornell. In it, he taught that technology is socially constructed—that what counts as a "working" technology is not given by its material properties but negotiated among social groups with different interests, different criteria of success, and different degrees of power. He taught that testing is never just technical. He taught, in his way, a form of ordinary hope: that if the meaning of a technology is negotiated rather than determined, then it is always possible—never guaranteed, but always possible—to negotiate differently.

Every call to a voice AI helpline is a negotiation. The caller brings a need; the system brings a database, a taxonomy, a voice. What happens between them is not predetermined by the code. It depends on whether the system can hear the caller, whether the database contains what the caller needs, whether the referral actually leads somewhere—and on whether someone, somewhere, has done the patient, invisible, unglamorous work of maintaining the infrastructure through which care flows.

The question Trevor would ask, I think, is not "Does this technology work?" but "Who decides whether it works?" And the answer—if we take SCOT seriously, if we take infrastructure studies seriously, if we take care ethics seriously—is that the people who should decide are the people who call. Not the designers, not the funders, not the city

officials, not the AI companies. The callers. The elderly woman whose heat has been shut off. The new arrival who does not know where to find food. The person leaving a dangerous situation who needs somewhere safe tonight.

They are the relevant social group. They are the testers. Their experience is the test.

This paper draws on the author's experience designing a voice AI system for community resource navigation in New York City and on the design methodology described in a companion paper (in preparation). It does not disclose system architecture, implementation details, or proprietary methods.

An earlier version of the argument about algorithmic temporality appears in Michalove and Jackson (2025). The author thanks Steven J. Jackson for supervision and for the hope.

References

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press, Cambridge, UK.

Bowker, G. C. and Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA.

Boyum, S., Kreuter, M. W., McQueen, A., Thompson, T., and Greer, R. (2016). Getting help after calling 2-1-1: A follow-up study of referral outcomes. *Journal of Social Service Research*, 42(3):279–292.

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness*. Duke University Press, Durham.

Crary, J. (2013). *24/7: Late Capitalism and the Ends of Sleep*. Verso, London.

Douglas, M. (1966). *Purity and Danger: An Analysis of the Concepts of Pollution and Taboo*. Routledge, London.

Downer, J. (2011). "737-cabriolet": The limits of knowledge and the sociology of inevitable failure. *American Journal of Sociology*, 117(3):725–762.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, New York.

Federal Communications Commission (2000). The use of N11 codes and other abbreviated dialing arrangements. Third Report and Order and Order on Reconsideration, FCC 00-256.

Fischer, C. S. (1992). *America Calling: A Social History of the Telephone to 1940*. University of California Press, Berkeley.

Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. Vintage Books, New York. Translated by Alan Sheridan. Originally published 1975.

Green, V. (2001). *Race on the Line: Gender, Labor, and Technology in the Bell System, 1880–1980*. Duke University Press, Durham.

Hochschild, A. R. (1983). *The Managed Heart: Commercialization of Human Feeling*. University of California Press, Berkeley.

Jackson, S. J. (2014). Rethinking repair. In Gillespie, T., Boczkowski, P. J., and Foot, K. A., editors, *Media Technologies: Essays on Communication, Materiality, and Society*, pages 221–240. MIT Press, Cambridge, MA.

Jackson, S. J. (2023). Ordinary hope. In *Ecological Reparation: Repair, Remediation and Resurgence in Social and Environmental Conflict*, pages 417–433. Cambridge University Press, Cambridge.

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Tober, C., Ricketts, C. R., Jurafsky, D., and Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14):7684–7689.

Larkin, B. (2013). The politics and poetics of infrastructure. *Annual Review of Anthropology*, 42:327–343.

Mattern, S. (2018). Maintenance and care. *Places Journal*.

Michalove, J. and Jackson, S. J. (2025). Algo-rhythmanalysis: Ethnographic methods for studying how algorithms structure time. Working draft. Pre-print available at https://silver.is/derive/.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, New York.

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*. Basic Books, New York.

Pinch, T. (1993). "testing—one, two, three . . . testing!": Toward a sociology of testing. *Science, Technology, & Human Values*, 18(1):25–41.

Pinch, T. J. and Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, 14(3):399–441.

Puig de la Bellacasa, M. (2011). Matters of care in technoscience: Assembling neglected things. *Social Studies of Science*, 41(1):85–106.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press, New Haven.

Shotwell, A. (2016). *Against Purity: Living Ethically in Compromised Times*. University of Minnesota Press, Minneapolis.

Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, 43(3):377–391.

Star, S. L. and Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7(1):111–134.

The Care Collective (2020). *The Care Manifesto: The Politics of Interdependence*. Verso, London.

Tracey, P. and Garcia, P. (2024). After automation: Homelessness prioritization algorithms and the future of care labor. *Big Data & Society*, 11(1).

Tronto, J. C. (1993). *Moral Boundaries: A Political Argument for an Ethic of Care*. Routledge, New York.

Tronto, J. C. (2013). *Caring Democracy: Markets, Equality, and Justice*. NYU Press, New York.

UNESCO (2019). I'd blush if i could: Closing gender divides in digital skills through education. Technical report, UNESCO, Paris.

Virilio, P. (2006). *Speed and Politics*. Semiotext(e), Los Angeles. Originally published 1977.

Wajcman, J. (2015). *Pressed for Time: The Acceleration of Life in Digital Capitalism*. University of Chicago Press, Chicago.

Woolgar, S. (1990). Configuring the user: The case of usability trials. *The Sociological Review*, 38(S1):58–99.