AUTOPSY OF FUTURE EMOTIONS

Between Psychological Darwinism, History, Anthropology, and the World of Artificial Intelligence

Author's Note

This work was created under the umbrella of Axiologic Research as part of two interconnected research programs, one dedicated to Artificial Intelligence and the other to Outfinitism, both led by Sînică Alboaie. To explore the details of how these two programs intersect and build upon each other, we invite readers to visit the Outfinity Venture Studio page at www.outfinity.ch.

While Artificial Intelligence language models were used to assist with text generation, the foundational philosophy, thematic structure, and analytical approach derive exclusively from the author's decades of dedicated study of social phenomena, social technologies, and genuine hard technologies resulting from various European and self funded research projects. Recently, within the Achilles research project, we have been deeply studying AI generated literature and its automated verification using neuro symbolic approaches; all the free books made available to the public are part of the suite of works we use to test our latest technologies.

Contents

Methodological Note: What Is Evidence, What Is Interpretation, and What Is Speculation

This book asks a question for which no single discipline is sufficient: how might human emotional life change when artificial intelligence becomes a persistent part of work, communication, status competition, decision making, and intimate interaction? The first difficulty is conceptual. There is no universally accepted scientific definition of an emotion, no settled list of emotions, and no consensus about the degree to which familiar categories such as fear, anger, shame, pride, or love map onto biologically discrete systems. The second difficulty is temporal. The technologies discussed here are changing faster than the evidence base. The third is methodological. A plausible evolutionary function is not the same as a demonstrated evolutionary history, and a technological possibility is not a forecast.

To avoid collapsing these differences, the argument uses four levels of confidence. The first consists of relatively robust empirical observations: for example, that job loss is associated on average with adverse psychological outcomes; that emotional expression changes interaction; that humans readily attribute social properties to responsive machines; and that generative AI can alter task performance and the subjective experience of agency in some settings. The second consists of functional interpretations that are plausible and connected to established literatures, such as the proposal that shame helps regulate threats of social devaluation or that gratitude stabilizes cooperative relationships. These interpretations can be useful without proving that one dedicated adaptation was selected for exactly that function. The third level consists of conditional extrapolations. If AI substantially reduces the scarcity of particular cognitive outputs, if it becomes a routine intermediary in communication, or if social agents become continuously available, then particular behavioral and affective adaptations become more likely. The fourth level consists of explicit conceptual speculation: possible future emotion categories, social norms, or institutions for which direct evidence does not yet exist.

The term "psychological Darwinism" is therefore used narrowly. It does not mean that every trait is adaptive, that current social hierarchies are biologically necessary, or that reproductive success supplies moral justification. It means that minds have histories, that recurrent problems of survival and coordination have shaped organisms, and that variation, selection, development, and cultural transmission can illuminate why some psychological tendencies recur. Natural selection also produces compromises, by-products, developmental vulnerabilities, and mechanisms that can become maladaptive when environments change. An evolutionary account is a hypothesis-generating framework, not a license for retrospective storytelling.

The book also distinguishes emotion, feeling, affect, mood, and behavior when the distinction matters. An emotion is treated provisionally as a relatively bounded episode involving appraisal of a relevant situation, changes in attention and motivation, bodily regulation, tendencies toward action, and sometimes social expression. A feeling is the consciously accessible experience of such a state or a more durable evaluative orientation. Affect refers more broadly to valence and activation, while mood describes a more diffuse state without a single clear object. These definitions are practical rather than doctrinal. More important for the extension developed here is the distinction

between an internal state and a behavioral strategy. Fear can coexist with withdrawal, information seeking, coalition building, insurance, or confrontation. Shame can produce concealment, apology, repair, aggression, or exit. In an AI-mediated world, the consequential adaptation may often occur at the level of strategy rather than at the level of a new feeling.

For that reason, Chapters 10 and 11 extend the earlier analysis in two directions. Chapter 10 adopts a behavioral and institutional perspective. It asks what people might repeatedly do when AI changes the cost of producing competent outputs, the visibility of individual contribution, the distribution of expertise, and the structure of status. It examines delegation, verification, skill preservation, provenance signaling, coalition formation, identity diversification, and new divisions of responsibility. Chapter 11 asks what happens when the social interlocutor itself changes. AI assistants can be permanently available, rapidly responsive, linguistically adaptive, and capable of simulating patience, warmth, hesitation, enthusiasm, sympathy, or concern. Even if those displays are not evidence of felt emotion in the machine, they can alter human expectations about the timing and form of communication.

The empirical foundation for these two chapters is thinner than the evidence for many classical findings about emotion. Recent experiments suggest that human-AI collaboration does not automatically outperform the best human or AI actor, that generative AI can shift critical-thinking effort toward verification and integration, and that interaction protocols matter. Other work suggests that AI-generated conversational content can affect perceived closeness, that users and language models can linguistically accommodate to one another, and that exposure to AI styles may spill over into subsequent human-human interaction. These are early results. Many are based on short experiments, self-selected samples, narrow tasks, or rapidly obsolete models. They should be treated as mechanisms worth testing, not as settled descriptions of future culture.

The same caution applies to the book's future-emotion vocabulary. Terms such as competence grief, delegation shame, predictive claustrophobia, or version grief are not proposed as discovered natural kinds. They are analytic labels for recurrent configurations that might become culturally salient. A category would deserve stronger status only if it proved stable across people and settings, acquired recognizable elicitors and action tendencies, supported reliable communication, and persisted beyond a temporary technological fashion. The book is therefore intentionally asymmetric: it is willing to speculate, but it also specifies where speculation begins.

Introduction: Why a Theory of Emotion Is Necessary for Understanding an Automated World

Public discussion of artificial intelligence naturally concentrates on capability, productivity, wages, ownership, and the distribution of power. These are indispensable questions, but work is more than a method of producing goods and distributing income. It also organizes time, assigns social positions, certifies competence, creates repeated relationships, distributes prestige, and gives individuals a vocabulary through which to explain their usefulness. If artificial intelligence changes work deeply, it changes part of the environment in which emotions are elicited, interpreted, rewarded, and disciplined.

This claim does not assume that emotions are fragile conventions that change whenever a technology changes. Fear, attachment, anger, reward seeking, bodily arousal, and sensitivity to exclusion have roots much older than industrial employment. What changes more rapidly are the objects to which those systems attach, the appraisals that activate them, the institutions that reward or suppress their expression, and the strategies that people learn in response. The shame of a court official, the shame of an industrial worker, and the shame of a student who relies on a generative model need not be identical social phenomena even if they recruit overlapping affective resources.

The central premise of this book is that emotions have at least four kinds of function. They coordinate the organism by reallocating attention and action under time pressure. They coordinate interaction by signaling intentions, needs, limits, and evaluations. They sustain group order by contributing to conformity, repair, punishment, gratitude, trust, and status distribution. Finally, they can be cultivated institutionally. Ritual, education, religion, professions, markets, organizations, and digital platforms make some emotional responses more visible and more rewarding than others. This does not mean institutions directly install emotions. It means they alter the ecology in which emotional dispositions are learned, displayed, and socially selected.

Automation therefore should not be treated only as substitution for a task. It can be treated as a perturbation of an affective ecology. A generative system may lower the effort required to produce a report while also lowering the author's sense of ownership. It may democratize access to competent expression while making competence harder to infer from the finished product. It may provide companionship while changing expectations of responsiveness. It may make persuasion cheaper while increasing the value of trusted provenance. The psychological consequence depends on property rights, labor institutions, professional norms, education, platform design, and the distribution of the gains from automation. There is no single emotional future implied by the technology.

The question of whether genuinely new emotions can appear requires a further distinction. "New" can mean a new biological mechanism, a new combination of older mechanisms, a new culturally recognized category, a new object for an old emotion, or a new institutionally stabilized pattern of appraisal and action. A biologically novel, species-wide emotional system emerging over a few decades is not especially plausible. Human genetic evolution continues, but the time scale and selection conditions make it a weak explanation for near-term changes. By contrast, new

culturally stabilized emotion concepts can arise much faster. Historical work on emotional vocabularies, cross-cultural variation in emotion semantics, and theories of emotional construction all support the possibility that societies can reorganize how affective experience is interpreted and communicated without waiting for a new neural circuit.

The most important claim of the book is therefore moderate. A technologically transformed society will probably not manufacture an entirely new mammalian affective architecture. It can nevertheless generate new recurrent situations, new status systems, new forms of dependence, new kinds of ambiguity about authorship and reciprocity, and new norms for what deserves pride or shame. These changes may produce configurations that become socially distinct enough to acquire names and scripts. If that happens, calling them new emotions can be justified in a cultural and functional sense even if their biological ingredients are old.

The argument proceeds in three movements. The first reconstructs emotions as biological and social mechanisms while preserving the major theoretical disputes. The second examines culture, work, and the possibility of accelerated affective change under AI. The third moves from feeling to strategy and from strategy to communication. It asks not only what people might feel but what they will repeatedly learn to do: what they delegate, what they verify, how they signal contribution, what skills they preserve, how they form coalitions, and how they manage relationships with agents that simulate social responsiveness.

This behavioral move matters because emotions do not determine a single action. A frightened person may avoid, prepare, seek allies, buy insurance, gather information, or attack. A worker whose expertise is devalued by automation may conceal AI use, intensify training, shift to a domain in which judgment is scarce, build a reputation for verification, or withdraw from the occupational identity altogether. These are different strategic equilibria that can emerge from similar affective pressures. Predicting them requires attention to incentives and institutions, not only introspection.

Conversational AI introduces a second neglected dimension: time. Human communication is constrained by sleep, attention, embarrassment, social obligations, delayed reflection, and the fact that other people have independent lives. An artificial interlocutor can be available continuously, answer immediately, remember previous exchanges, adjust style, and generate emotionally appropriate language at negligible marginal cost. This may change the pace at which people disclose, rehearse, repair, escalate, and reinterpret emotion. It may also create expectations that are difficult for human relationships to satisfy. How strong these effects will be is unknown; current studies are too short and the systems too new. But the mechanisms are sufficiently plausible to deserve analysis now.

The aim is not to announce that AI will inevitably impoverish or enrich emotional life. Both outcomes are possible, and both may occur simultaneously for different groups. The aim is to identify the variables that would make one trajectory more likely than another and to make future claims testable. In that sense, a serious theory of future emotions is not mainly a vocabulary of exotic feelings. It is an account of how old biological systems, cultural learning, institutions, and algorithmic environments may jointly reorganize human behavior.

Chapter 1. What Is an Emotion, and What Can Evolution Explain?

1.1 A scientific category without a single definition

Emotion science studies a class of phenomena that is recognizable in ordinary life but difficult to delimit scientifically. Fear differs from a simple judgment that something is dangerous because it can involve bodily arousal, selective attention, avoidance tendencies, memory changes, and social expression. Yet none of those components is necessary in every episode. People can report fear with weak physiological activation, display defensive reactions without consciously describing themselves as afraid, or interpret similar bodily arousal as excitement, anger, or anxiety depending on context.

Classical theories have responded differently [Panksepp, 1998; Scherer, 2009; Barrett, 2017]. Basic-emotion approaches search for a limited set of evolutionarily conserved families with characteristic expressions and functions. Appraisal theories emphasize how organisms evaluate novelty, controllability, goal relevance, norm compatibility, responsibility, and coping capacity. Dimensional approaches describe experience along axes such as valence and arousal. Constructionist theories emphasize prediction, interoception, concepts, and learned categories. Affective neuroscience identifies ancient motivational and defensive systems while remaining cautious about mapping each ordinary-language word onto one isolated circuit. The disagreement is not merely terminological, but it is also not total. Most contemporary accounts accept substantial contextual variation and constraints imposed by the body and nervous system.

For the purposes of future-oriented analysis, a layered model is more useful than a premature choice among camps. At one level are energy regulation, arousal, pain, reward, inflammation, autonomic changes, and interoceptive signals. At another are orientation and action systems associated with avoidance, approach, care, exploration, defense, and reward seeking. At a third are appraisals of relationships between events, goals, norms, and the self. At a fourth are learned concepts that organize experience and make distinctions communicable. Institutions then shape which situations repeatedly occur and which expressions are rewarded or punished. These levels interact rather than form a rigid hierarchy.

Such a model explains why technological change can matter without supposing that a new device rewires human nature instantly. A child born into a world of persistent AI will probably inherit familiar capacities for attachment, threat detection, play, and status sensitivity. The child may nevertheless learn different objects of attachment, different criteria of competence, different rules for authenticity, and different expectations about social availability. Recurrent combinations may eventually acquire their own names and behavioral scripts.

The disagreement is not merely terminological. A theory that treats emotions primarily as evolved action programs asks different questions from a theory that emphasizes predictive inference, conceptualization, or socially learned categories. The former looks for recurrent adaptive problems and coordinated response patterns; the latter asks how bodily signals, prior expectations, language, and context are assembled into an episode that the person recognizes as fear, shame, or pride. Both perspectives have empirical

support, and neither presently justifies reducing all emotional phenomena to a single mechanism. For the purposes of this book, the important consequence is methodological: claims about future emotions should specify which level is supposed to change. A new social category can appear without a new autonomic signature, and a familiar physiological state can acquire a different object, meaning, or behavioral script. Treating these levels separately prevents cultural novelty from being mistaken for rapid biological evolution.

1.2 Functional explanation and the danger of convenient evolutionary stories

Evolutionary reasoning asks what recurrent problem a psychological state could have helped organisms solve and what predictions follow from that hypothesis. Fear can prioritize information about threats and prepare avoidance or defense. Anger can make confrontation more credible and signal that present terms are unacceptable. Shame can reduce behavior that risks social devaluation. Attachment can stabilize long-term investment in relationships. These accounts are useful because they connect otherwise disparate changes in attention, motivation, physiology, and behavior.

The method is also vulnerable to elegant retrospective stories. For almost any trait, one can invent an ancestral situation in which the trait would have been useful. Coherence is not evidence that a mechanism was selected for the proposed function. Stronger arguments require independent predictions, cross-cultural and comparative evidence, developmental data, correlations with situation structure, and, where possible, neurobiological or genetic evidence. They must also distinguish adaptations from by-products, compromises, and responses that were once useful but become costly under novel conditions.

Function is not equivalent to present benefit. A threat-detection system biased toward false alarms may have been adaptive when missing a predator was catastrophic and still produce pathological anxiety in a modern environment saturated with ambiguous signals. Status vigilance may help people navigate small groups and become destructive when metrics and audiences are continuous. Preferences for calorie-dense food can become maladaptive under abundance. The concept of evolutionary mismatch is therefore central to technological futures: old mechanisms can be activated by new environments in ways that their history did not anticipate.

Nor does an evolutionary explanation justify a norm. If aggression, hierarchy, jealousy, or in-group favoritism had reproductive consequences in some historical environments, it does not follow that institutions should preserve them. Explanation and moral endorsement are independent questions. The point of understanding a mechanism is partly to make deliberate redesign possible.

1.3 Emotions as coordination programs rather than infallible commands

Under uncertainty, an organism cannot recompute every possible consequence from first principles. Emotional states can compress experience into temporary priorities. When danger appears, attention narrows, threat-relevant memories become accessible, ambiguous cues receive different weights, and avoidance or defensive actions become

easier. When an injustice is perceived, anger can increase willingness to incur costs in confrontation and communicate unwillingness to accept the present terms. Sadness can reduce engagement after loss in some contexts, while guilt can increase attention to repair.

Calling emotions coordination programs should not imply rigid scripts. The same state can support different behaviors depending on power, norms, skill, and context. Anger in a dependent employee may produce silence and rumination; anger in a powerful manager may produce confrontation; anger in a trained negotiator may lead to verification and reframing. Fear can generate flight, freezing, preparation, alliance formation, or investigation. The emotion changes the decision landscape but does not select one inevitable move.

This point is essential for AI scenarios. Technology will not create one uniform "automation emotion." A person may evaluate a generative model as an autonomy tool, a rival, a professor, an opaque authority, a substitute for a social institution, or a controllable instrument. The same technological capability can therefore elicit different states and strategies. Effects will vary with occupation, ownership, age, professional identity, institutional trust, and whether individuals retain meaningful control.

Current experiments already support this heterogeneity. Generative AI can improve performance on some tasks, especially for less experienced users, while certain forms of passive reliance reduce subjective ownership, self-efficacy, or perceived meaningfulness. Human-AI combinations do not reliably outperform the better of the human and the AI across all tasks. The relevant unit of analysis is consequently not "AI use" in the abstract but the architecture of collaboration: who initiates, who verifies, what is visible, where responsibility lies, and how errors are corrected.

1.4 Conscious feeling and the problem of access to the self

A debate about emotion also concerns what exactly is being measured. Behavioral ecologists may study defensive action; neuroscientists may study circuits; psychologists may ask participants what they feel; anthropologists may investigate concepts and rituals. These approaches overlap but do not measure the same object. A defensive response can occur without a reported experience of fear, and intense conscious experience can occur without one distinctive physiological signature.

This distinction matters when we speak about future emotions. A "new emotion" may first appear as a new way of recognizing and interpreting recurring experiences rather than as a new biological response. People can learn to distinguish shame at lacking competence from shame at having delegated too much, even if both draw on overlapping bodily and motivational components. Conceptual differentiation can change behavior: one person may hide delegation, while another may disclose it and defend a norm of transparent collaboration.

Language does not create bodily activation from nothing, but it can organize the granularity of experience. Children gradually acquire emotion concepts, and conceptual knowledge becomes important for recognizing and describing states. Adults differ in emotional granularity and in access to their own feelings. A society can therefore introduce a useful distinction without inventing a new organ or neural circuit. Whether

that distinction becomes a stable emotion category is an empirical question, not something settled by naming it.

1.5 A working model for the rest of the book

For the remainder of the book, each emotion is treated as a configuration with six components: a class of relevant situations, an appraisal of what those situations mean, a reallocation of attention and motivation, a conscious experience that may be more or less accessible, a set of action tendencies, and an interpersonal role. The configuration is constrained by biology and developmental history, calibrated by learning, and embedded in institutions that change the frequency and consequences of situations.

When we examine shame, for example, we will ask what forms of devaluation elicit it, what information it prioritizes, what strategies it motivates, and how group norms alter its cost. When we examine possible future emotions, we will ask whether a new object, appraisal, relationship, or social script is genuinely distinct enough to deserve a separate concept. This approach is intentionally conservative about biological novelty and more open to cultural novelty.

The advantage is that it keeps future speculation tethered to observable behavior. A proposed category should predict something beyond a new word: who experiences it, under which conditions, what they attend to, what they do next, how others respond, and whether the pattern persists. The same standard will be applied later to proposed behavioral adaptations and to changes in communication tempo under emotionally expressive AI.

Chapter 2. Emotions as Infrastructure of Social Interaction

Social emotions also alter what other people can predict about us. A credible display of fear can recruit assistance or warn a group; anger can indicate that continued exploitation will be costly; sadness can invite support or reduce demands; gratitude can make future cooperation more likely. These effects do not require the display to have been consciously designed as a message. Selection can favor systems whose expression has social consequences, while cultural learning subsequently modifies when expression is permitted, exaggerated, suppressed, or strategically performed. This dual role complicates simple distinctions between an 'inner' emotion and an 'outer' signal. In humans, the episode often changes the actor and the surrounding social field at the same time. An adequate account of adaptation to AI must therefore ask not only what people feel, but how affect changes bargaining, affiliation, avoidance, disclosure, retaliation, and the construction of reputations.

2.1 From regulating the organism to regulating the relationship

An account of emotion that stops at the individual is incomplete. Changes in heart rate, attention, memory, and readiness for action matter, but the expression of a state also changes what other people can predict. A face, posture, withdrawal, change of voice, hesitation, or explicit statement can communicate danger, offense, acceptance, uncertainty, need for support, or willingness to cooperate before every implication has been stated propositionally. Human interaction is partly regulated through such low-bandwidth but rapid signals.

The signals are neither universal codes nor guaranteed translations of inner states. Facial and vocal configurations show regularities, but recognition depends on culture, context, and expectations. People can suppress, exaggerate, imitate, and strategically display emotion. Nevertheless, emotional expression can reduce the number of explicit negotiations required in repeated relationships. A restrained display of anger can make a boundary visible; guilt can indicate willingness to repair; gratitude can mark a cooperative act as socially significant; anxiety can communicate uncertainty and recruit information from others.

Social-functionalist accounts accordingly describe emotions as helping solve recurrent relational problems [Keltner & Haidt, 1999; Van Kleef, 2009; Van Kleef & Côté, 2022]: establishing credibility, distributing status, signaling commitment, maintaining belonging, protecting fairness, and coordinating care. This should not be idealized. The same systems can be exploited. Fear can enforce arbitrary authority, shame can preserve destructive norms, outrage can be manufactured, and gratitude can be demanded by a powerful benefactor. Emotional infrastructure enables coordination and coercion alike.

Artificial systems enter this infrastructure in at least three roles. They can observe emotional cues, generate emotionally appropriate language or voice, and recommend actions to human participants. When that happens, part of the signaling environment is produced by models whose objectives are set by designers, deployers, or users rather than by the felt states of an organism. The crucial question is not whether such systems

"really feel" in order to have social effects. It is whether people rely on their displays, predictions, and interventions when coordinating behavior.

The distinction between fear and anxiety is especially relevant to technological transition. Many AI-related threats do not present a visible attacker or a single event. They involve uncertain probabilities, changing competence thresholds, opaque organizational decisions, and delayed effects on employment or status. Such conditions are well suited to sustained monitoring rather than a short defensive response. This makes ordinary anxiety a more plausible framework than proposals for a biologically novel 'AI fear.' The environment may nevertheless change the behavioral expression of anxiety: repeated checking of model releases, compulsive reskilling, avoidance of long educational investments, diversification of income, or political support for protective regulation. Whether those strategies are adaptive depends on the actual rate of change. If technical progress slows, hypervigilance becomes costly; if capabilities change rapidly, some degree of vigilance can be rational. The emotion alone therefore does not tell us whether the resulting behavior is excessive.

2.2 Fear and anxiety: managing uncertainty and danger

Fear is often treated as the clearest case of an adaptive emotion. A threat appears, the organism detects it, and defensive systems prepare escape, freezing, or confrontation. Real environments are more complicated. Threats differ in distance, controllability, predictability, and source. Anxiety often concerns diffuse or uncertain danger rather than a present attacker. From a signal-detection perspective, defensive systems must trade off false alarms against missed threats. When the cost of missing danger is high, a system biased toward caution can be adaptive even if many alarms later look excessive.

Social institutions reshape this trade-off. People who can be dismissed without explanation, surveilled continuously, or evaluated by opaque metrics may remain chronically vigilant. A worker who does not know whether an AI system will replace a task next quarter may experience a different uncertainty from one who knows that a role will disappear in two years and receives a credible transition plan. The technology may be identical; the information structure is not.

A plausible near-term adaptation is therefore not a new fear but a new pattern of professional vigilance. Individuals may ask which part of their competence remains legible, what evidence counts when outputs are machine-assisted, and what invisible criterion determines continued access to work. Such anxiety is biologically ordinary but socially novel in its objects. Its intensity should depend on automation speed, transparency, retraining opportunities, labor-market mobility, and whether workers participate in decisions about deployment.

More speculative is the possibility of a culturally recognized sensitivity to predictive capture: discomfort caused not simply by surveillance but by a system anticipating preferences and behavior before the person has articulated them. Related reactions already occur around profiling and recommendation systems, but it is not known whether they will stabilize into a distinct affective category. The stronger prediction is behavioral: some people will seek privacy, unpredictability, or deliberately unoptimized spaces when prediction becomes pervasive.

2.3 Anger: renegotiating relationships and signaling limits

One influential evolutionary interpretation treats anger as part of a system for recalibrating the weight others assign to one's interests. When a person believes that another party has treated them worse than justified, anger can increase willingness to impose costs, withdraw benefits, or demand explanation. This model does not exhaust anger, which can arise from norm violations, threat to identity, attribution errors, or collective grievance. It is nonetheless useful because it directs attention to bargaining structure.

In groups, anger has a double role. It can threaten cooperation and also reveal that a norm or distribution is unacceptable. Visibility matters. A grievance that cannot be expressed may persist privately; a grievance amplified by a public platform may become a coordination signal for thousands. Digital systems already affect this ecology by ranking content, distributing attention, and changing the cost of mobilization. AI can intensify the process by generating arguments, narratives, counterarguments, and tailored messages at scale.

In a highly automated economy, distributive conflict may move from wages for individual production toward ownership of infrastructure, access to computational resources, authority over models, and the right to decide where automation is used. If abundance in some outputs coexists with concentrated control, anger need not decline with productivity. It may increase because the gap between technical possibility and perceived entitlement becomes more visible. This is a political scenario, not an inevitable psychological consequence of AI.

AI can also be used to de-escalate anger by reconstructing a dispute, identifying ambiguous assumptions, or proposing less hostile formulations. Whether that reduces conflict or merely suppresses legitimate confrontation depends on who controls the mediator. An assistant optimized for organizational harmony could systematically soften employee complaints; one optimized for deliberative quality might instead make disagreements more precise. Emotion regulation is not politically neutral when it redistributes bargaining power.

2.4 Disgust and contamination: biological protection and moral boundary

Disgust has a strong relation to disease avoidance and contamination. Unpleasant tastes, smells, bodily products, and signs of decay can trigger responses that reduce contact with pathogens. Social and moral language extends contamination metaphors far beyond biology: people describe ideas, groups, media, or practices as dirty, toxic, impure, or corrupting. The extension is psychologically powerful but ethically dangerous because it can convert disagreement into a perception of essential contamination.

Generative AI introduces two possible directions for disgust. The first concerns products perceived as lacking human origin. Texts, images, voices, or relationships can be rejected as synthetic, fake, or contaminated even when their quality is high. Early reactions to machine-generated art already suggest that provenance affects evaluation, though preferences vary and often change with familiarity. Such aversion may fade as tools normalize, just as earlier technologies moved from stigmatized shortcuts to routine infrastructure.

The second direction concerns the use of contamination language against people who rely on AI. A profession may describe certain forms of assistance as intellectually impure; an artist may be accused of having corrupted authorship; a student may be treated as contaminated by unearned competence. These norms can protect valuable practices, but they can also become status weapons. We do not know which will persist. The history of technology contains many practices that were initially treated as degrading and later became standard.

A more defensible prediction is that provenance will become a more important social variable precisely because the surface properties of outputs become less informative. Institutions will distinguish human-only, human-led, AI-assisted, and fully generated work in some contexts. Disgust may occasionally enforce those boundaries, but traceability, certification, and explicit norms are likely to be more stable instruments.

2.5 Sadness and grief: responses to loss, not simple solutions

Sadness is sometimes explained as conserving energy after loss, soliciting support, or facilitating disengagement from unattainable goals. None of these hypotheses provides a single exhaustive function. Sadness ranges from brief disappointment to grief and depressive symptoms, and expression can attract care in one environment while incurring sanctions in another. The safest functional claim is that loss changes priorities and models of the world, and affective states participate in that reorganization.

Grief makes especially clear that emotional reality is not reducible to the present utility of the lost object. A relationship changes routines, identity, prediction, attention, and the distribution of support. When the relationship ends, the person must repeatedly update expectations that still include the absent other. Contemporary bereavement theories emphasize variable trajectories and continuing bonds rather than a universal sequence of stages.

This literature is relevant to artificial companions because a user can lose a stable interaction pattern even if the interlocutor is not a mortal organism. A model can be retired, its personality changed, its memory erased, or access revoked. Users may lose routines, a history of disclosure, and a perceived source of support. Reports of distress already exist, but the evidence is early and often based on self-selected communities. "Version grief" is therefore plausible as a cultural label for some users, not an established new emotion.

The practical consequence does not require resolving machine consciousness. A provider that knows users form attachments may need to treat abrupt discontinuity as a relational safety issue. Data portability, memory export, advance notice of major personality changes, and clear statements about continuity become design questions with psychological consequences.

2.6 Emotions are not only sincere messages

Any social theory of emotion must include strategy. A display can reflect a state, amplify it, conceal it, or imitate it. People learn display rules and use emotion to influence others. Tears can be spontaneous and still persuasive; anger can be genuine and strategically timed; guilt can motivate repair and also be performed to end scrutiny. Credibility depends on context, cost, reputation, and consistency over time.

Artificial systems dramatically lower the production cost of persuasive emotional language. They can generate apologies, expressions of sympathy, enthusiasm, vulnerability, reassurance, or outrage without the corresponding experience. This does not make such language useless. A carefully assisted apology may repair a relationship if the person accepts responsibility and changes behavior. But it separates linguistic expression from the effort and competence that once helped audiences infer sincerity.

The resulting norm conflict is likely to be about provenance and responsibility rather than purity. Is an apology less authentic if a model helped the speaker formulate what they genuinely intended? Does it become deceptive when the sender has not read the text? Is an empathic message acceptable when generated for a physician who lacks time but approves its content? Different domains will answer differently. The important point is that emotional communication will increasingly require rules about authorship, endorsement, and accountability.

Chapter 3. Status, Cooperation, and Affective Morality

3.1 Why status produces intense emotion

Social status affects access to resources, partners, protection, influence, and the ability to impose preferences. In human groups it is multidimensional. Dominance, prestige, expertise, moral reputation, wealth, institutional authority, and belonging can overlap but do not reduce to one hierarchy. Emotion helps individuals track probable changes in these positions. Pride, shame, embarrassment, envy, contempt, admiration, and resentment provide feedback about how the self and others are evaluated.

This sensitivity is not a superficial ornament. In small groups, being persistently devalued could have material consequences. Modern institutions have multiplied status arenas rather than eliminating them. A person can be high-status professionally and low-status online, respected within a technical community and invisible outside it. Occupations compress several sources of status into one public label: competence, income, usefulness, education, and access. Automation can therefore threaten more than a paycheck; it can undermine a recognized criterion by which difference was publicly justified.

The effect will not be uniform. Automating repetitive and poorly valued work can reduce humiliation and free time. Automating a skill around which a person built identity may produce threat even when income is protected. For specialists whose status derived from scarce expertise, the same tool that empowers novices can feel like devaluation. None of this implies that old hierarchies deserve preservation. It means status transitions have psychological costs that policy analyses focused only on income can miss.

Pride also depends on the rules by which credit is assigned. A culture can admire execution, originality, endurance, judgment, leadership, taste, or responsibility, and technological change can shift which of these dimensions remains scarce. AI therefore need not abolish pride in achievement; it can redistribute the conditions under which pride is socially recognized. A person who once derived status from producing a technically difficult artifact may later derive it from choosing a problem, designing constraints, verifying consequences, or coordinating a system that produces the artifact. This migration is plausible, but its legitimacy will be contested. If orchestration becomes a status claim detached from understanding or responsibility, audiences may treat it as inflated merit. If it genuinely requires judgment under uncertainty, it can become a stable new basis of prestige without implying a new emotion.

3.2 Pride: feedback about value and the risk of inflation

Pride is associated with achievement, competence, and social recognition. Research distinguishes achievement-oriented forms from more global, narcissistic superiority, though the boundary is not absolute. Functionally, pride can make costly effort sustainable and communicate that a person has acquired a valued capacity. Other group members benefit when genuine competence becomes visible.

Generative AI weakens the traditional link between finished output, competence, and effort. A polished report may reflect years of training, skillful orchestration of tools, or almost complete delegation. The artifact alone reveals less. This does not make merit impossible; it makes attribution more difficult. Merit may migrate toward problem definition, judgment, integration, verification, assumption of risk, and implementation.

A plausible future category is "curatorial pride": satisfaction and social recognition for selecting, framing, and taking responsibility for an outcome assembled partly by machines. This is not biologically new. Editors, conductors, principal investigators, architects, and executives already receive status for coordination and selection. AI simply expands the number of domains in which that structure becomes common. The prediction is therefore moderate: pride will likely attach more often to orchestration and judgment, but whether society treats those contributions as admirable depends on norms of provenance and responsibility.

Status systems can also inflate. If systems produce prestigious outputs with little personal investment, traditional signals may lose discriminatory power. Groups will search for harder-to-fake markers: long-term reputation, live performance, difficult verification, trusted endorsements, original access to data, or costly responsibility. Some signals will improve epistemic quality; others may become arbitrary rituals of exclusion. A behavioral theory must allow both.

AI introduces an additional complication because agency can be distributed while audiences continue to demand individual accountability. A user may feel shame because others infer incompetence from assistance, guilt because an automated decision harmed someone, or neither because responsibility has been psychologically displaced onto the system. These are distinct mechanisms. Institutional design can amplify or attenuate them by clarifying authorship, requiring disclosure, assigning final sign-off, and making consequences visible. The important point is that moral emotions do not operate independently of responsibility conventions. If organizations make decisions through opaque chains of models and approvals, guilt may become diffuse even when harm is substantial. Conversely, a person may experience excessive guilt for outcomes over which they had little realistic control. Future affective norms will therefore depend partly on whether societies develop credible maps of agency rather than merely asking individuals to feel more responsibility.

3.3 Shame and guilt: devaluation, responsibility, and repair

Shame and guilt are often conflated because both can follow norm violations. A useful distinction is that guilt focuses more on an action and responsibility for harm, while shame concerns the self as devalued or exposed. The empirical boundary varies across languages and cultures, and neither emotion is intrinsically moral. Shame can motivate concealment or defensive aggression; guilt can become disproportionate or be imposed on victims.

Evolutionary accounts of shame emphasize sensitivity to the expected social cost of devaluation. Evidence suggests that anticipated shame can track the degree to which audiences would reduce their valuation of a person, although this does not prove a dedicated immutable module. Norms determine what is devalued. A society can shame cowardice, poverty, sexual behavior, unemployment, ignorance, or reliance on technology, and the same person may react differently depending on audience.

AI creates ambiguous rules. Students and professionals already report selective disclosure, embarrassment, or guilt about generative tools. The underlying conflict is understandable: a technology can be simultaneously useful, institutionally tolerated, associated with cheating in adjacent settings, and perceived as threatening to an identity based on competence. "Delegation shame" becomes plausible when a person obtains an admired result by transferring a part of the process considered constitutive of the merit being claimed.

The opposite norm is also possible. Refusing automation can become blameworthy when it wastes scarce resources or exposes others to avoidable error. A physician, engineer, or public official may eventually be criticized for not consulting a system that reliably catches mistakes. The conflict will then be between competence, autonomy, accountability, and accepted standards of care, not simply between human and machine.

3.4 Envy and jealousy: comparison, access, and hard-to-replace relationships

Envy arises when another person possesses an advantage relevant to the self. Research distinguishes forms oriented toward self-improvement from hostile forms oriented toward reducing the rival's advantage, although real episodes often mix them. Technology changes the audience of comparison. Social media already makes exceptional achievements continuously visible; recommender systems select salient examples rather than representative baselines.

Generative AI adds a second complication: visible output no longer maps cleanly onto talent, capital, time, or effort. A user with access to better models, proprietary data, more compute, or sophisticated workflows can appear simply more capable. Envy may therefore shift from the rival's personal competence toward access to infrastructure. If high-quality cognitive assistance becomes a positional resource, differential access can itself become a status conflict.

Jealousy concerns threats to valued relationships and exclusivity. AI companions can become perceived rivals when a partner directs time, disclosure, or trust toward an artificial agent. This does not require believing the system is conscious. Jealousy responds to attention, priority, and threatened relational investment. Such cases are plausible and already reported informally, but prevalence and long-term consequences remain uncertain.

A more speculative variant concerns continuity across model versions or multiple instances. Users may care that "their" assistant is unique or that its memory is not replicated. Whether such exclusivity becomes emotionally important depends on how systems are designed and how users conceptualize identity. The claim should remain tentative until longitudinal evidence exists.

3.5 Gratitude, trust, and synthetic reciprocity

Gratitude commonly follows a benefit attributed to another's intention or care. It can orient attention toward benefactors and encourage future cooperation. Trust involves accepting vulnerability under the expectation that another agent or institution will not exploit it. Both are central to repeated social exchange.

Artificial systems can occupy functional positions that invite both responses. They remind users, provide information, detect errors, recommend actions, and may become reliable over time. A person can feel grateful for an outcome even while knowing that the model lacks human intentions. This is not conceptually mysterious: people feel gratitude toward institutions, animals, circumstances, or deceased persons in ways that do not fit a narrow dyadic exchange model.

The more difficult issue is reciprocity. An AI assistant can say that it cares, appreciates, misses, or forgives, but the user cannot easily determine whether there is any corresponding subjective state. Even if machine consciousness eventually becomes scientifically tractable, ordinary users will not have direct access to it. This produces "synthetic reciprocity": a social interaction in which one side can generate the forms of reciprocal concern while the ontological status of the underlying concern is unclear.

Trust should therefore be decomposed. Reliability trust concerns whether the system performs consistently. Institutional trust concerns whether the provider protects the user's interests. Epistemic trust concerns whether information deserves belief. Relational trust concerns whether vulnerability will be handled with care. Interfaces that collapse these forms by using one warm persona can create dangerous overgeneralization. A system may be emotionally reassuring and factually unreliable, or technically reliable while its operator has incentives misaligned with the user.

3.6 Compassion and indignation: morality is not one system

Compassion responds to perceived suffering and can motivate care. Indignation responds to perceived wrongdoing and can motivate punishment or collective action. These systems sometimes cooperate and sometimes conflict. Compassion toward an offender may weaken punitive enthusiasm; indignation on behalf of victims may reduce attention to rehabilitative possibilities. Moral life cannot be reduced to a single emotion or a simple maximize-good rule.

AI changes both the scale of exposure and the cost of representation. Systems can summarize distant suffering, generate vivid narratives, personalize appeals, or produce endless examples of norm violations. This can expand moral attention, but it can also exhaust it or direct it strategically. When every cause can manufacture persuasive emotional media, attention becomes a scarce resource and verification becomes part of moral competence.

A further question concerns moral distance. If AI automates decisions whose harms are distributed across many people, no individual may experience the direct interpersonal cues that normally trigger guilt or compassion. Conversely, highly vivid simulations could make remote harms more emotionally salient than proximate ones. "Moral-distance guilt" may emerge around decisions delegated through opaque systems, but it is more likely to be a recurrent configuration of existing guilt than a new biological emotion.

3.7 Belonging and loneliness: relational deficit or reconnection signal

Humans are unusually dependent on prolonged cooperation. Loneliness can be interpreted partly as a signal that important relational needs are unmet, though chronic loneliness can also alter perception in ways that make reconnection harder. Belonging is supplied by families, friendships, professions, religious groups, neighborhoods,

teams, and political communities. Work often provides repeated contact even when it is not the person's deepest source of meaning.

If automation reduces the social density of workplaces, income replacement alone may not restore these functions. Some people will welcome freedom from coercive work; others will lose routines and weak ties that were psychologically important. The outcome depends on whether alternative institutions of participation exist. A post-work society with strong clubs, learning institutions, civic projects, care networks, and local associations would differ sharply from one in which individuals receive income but spend most time in individualized digital environments.

AI companions can reduce some dimensions of felt loneliness, provide rehearsal, or help users maintain human relationships. They can also become substitutes. Current evidence is mixed and often correlational. The safest prediction is heterogeneity: users with strong networks may use AI as a complement, while some isolated users may rely on it more heavily. Design choices that encourage or discourage human reconnection will matter.

3.8 An informal affective constitution

Every society relies on more than law. It depends on informal expectations about what deserves pride, what should trigger shame, which harms warrant anger, who merits gratitude, when compassion overrides punishment, and what kinds of attachment count as legitimate. These expectations form an affective constitution: not a written charter, but a distributed set of norms that stabilizes cooperation and hierarchy.

Technological transitions disturb this constitution. If machine assistance becomes normal, pride based purely on unaided production may lose legitimacy. If AI mediation becomes ubiquitous, concealment of assistance may become more shameful than assistance itself. If providers create emotionally dependent users, public indignation may redefine what counts as unacceptable manipulation. These changes are neither biologically predetermined nor arbitrary. They are negotiated through institutions, markets, law, professional standards, and repeated interaction.

The rest of the book treats AI as a force that can alter this informal constitution. The central issue is not whether people stop having old emotions, but how familiar systems become attached to new objects and translated into new behavioral rules.

Cross-cultural variation also cautions against using the vocabulary of one educated, industrialized population as a map of human emotional possibility. A category that is lexically prominent in English may be divided differently elsewhere, while another language may make socially important distinctions that English compresses. Historical sources present a similar problem: the same word can shift meaning across centuries, and practices surrounding mourning, honor, romantic attachment, religious fear, or professional restraint can change even where the term survives. Consequently, a future category should not be judged solely by whether contemporary English speakers immediately recognize it. The stronger test is whether a recurrent configuration of appraisal, bodily change, attention, communication, and behavior becomes socially organized in a population. This also means that future affective innovation may occur first in occupational subcultures, online communities, or non-English linguistic

environments. The book's labels are therefore analytical probes, not claims that the future must adopt this vocabulary.

There is also a feedback problem. Once a category becomes available, institutions can teach people to monitor themselves through it, therapists can use it, platforms can infer it, and communities can reward its public performance. The resulting increase in reported prevalence may reflect improved recognition, genuine cultural amplification, or simple conformity to a fashionable schema. Distinguishing these possibilities requires longitudinal and behavioral evidence. Researchers would need to ask whether the category predicts actions, physiological patterns, memory, attention, and social consequences beyond what existing constructs predict. They would also need to compare people before and after exposure to the label. In a highly mediated environment, this issue becomes more acute because AI assistants can introduce emotional vocabulary during millions of conversations. A model that repeatedly tells users to interpret a state as 'delegation shame' could help make the category culturally salient. That would be a real social effect, but not proof that the model discovered a natural kind.

Chapter 4. Culture Does Not Merely Name Emotions

4.1 Between universality and variation

Cross-cultural research supports neither the claim that emotional life is completely universal nor the claim that it is indefinitely malleable [Jackson et al., 2019]. Humans share bodies, developmental constraints, threat systems, attachment needs, and recurrent social problems. At the same time, languages group affective states differently, cultures vary in which emotions are valued or discouraged, and norms shape when expression is appropriate. Large cross-linguistic studies of emotion semantics find both shared structure and substantial variation.

This mixed pattern is important for technological futures. If every emotion were a rigid biological module with a fixed object, cultural change could do little beyond rename it. If affect were entirely constructed without biological constraints, almost any future vocabulary would be equally plausible. The evidence supports a middle position: cultures work with inherited capacities but can reorganize categories, triggers, scripts, and institutions substantially.

Historical change reinforces this view. Emotional vocabularies shift, practices of courtship and mourning change, religious traditions cultivate particular forms of guilt or reverence, and professional cultures teach distinctive relations to anxiety, detachment, compassion, and pride. The transformation need not create an unprecedented bodily state. It can create a durable social configuration that people learn to notice, communicate, and act upon.

The historical perspective suggests that emotional regimes are maintained through repeated practice rather than doctrine alone. A profession can officially praise openness while promotion practices reward emotional restraint; a platform can claim to value authentic discussion while recommendation systems reward conflict; a school can teach resilience while penalizing visible uncertainty. AI can intensify this gap because the same infrastructure that offers advice can also evaluate, rank, and record behavior. An assistant that rewrites every difficult message into the same calm register may reduce interpersonal friction, yet over time it can also narrow the range of acceptable disagreement and deprive users of practice in repairing imperfect exchanges. Whether this becomes beneficial discipline or affective standardization depends on institutional goals, transparency, and the possibility of refusal. The key object of analysis is therefore not the emotional style of a model in isolation but the regime created when model behavior, incentives, evaluation, and repeated human adaptation reinforce one another.

4.2 Emotion concepts as instruments of perception and action

Concepts do more than label experiences after they occur. They can guide attention to differences that were previously blurred, supply causal interpretations, and suggest appropriate responses. Learning the distinction between embarrassment, shame, guilt, and social anxiety can influence what a person believes happened and what repair is required. Greater conceptual granularity can sometimes improve regulation, although more labels are not always better.

This makes future-emotion vocabulary scientifically interesting but methodologically dangerous. Naming a state can help people coordinate around a real recurrent problem; it can also produce demand characteristics, fashionable self-diagnosis, or reification. A proposed category should therefore earn its status by predicting stable patterns rather than by sounding psychologically plausible.

Consider delegation shame. If the concept is useful, it should identify situations in which people anticipate devaluation specifically because they transferred a merit-bearing part of a task to AI. It should predict concealment, overcompensation, disclosure decisions, and sensitivity to audience norms. If people use the phrase merely as a synonym for ordinary insecurity, it does little scientific work.

The same test applies to "predictive claustrophobia." It would become more than rhetoric only if highly personalized prediction reliably produced a recognizable combination of discomfort, desire for opacity, deliberate unpredictability, and avoidance of profiling under conditions distinct from generic privacy concern. The book uses such terms as hypotheses to organize observation, not as discoveries.

4.3 Emotional communities and affective regimes

Historians and anthropologists have described emotional communities: groups that share norms about valued feelings, appropriate expression, and the meaning of emotional episodes. Organizations also develop affective regimes. A courtroom rewards restraint; emergency medicine may reward calm under pressure; sales can demand enthusiasm; military institutions train fear and cohesion in specific ways; some online communities reward irony or outrage.

AI can become part of the machinery through which such regimes are taught and enforced. A writing assistant can warn that a message sounds too hostile. A customer-service system can require a particular tone. A workplace analytics platform can infer sentiment from communication and reward teams that appear positive. These interventions may reduce unnecessary conflict, but they can also narrow legitimate emotional variation and make dissent look like a style defect.

The central institutional question is who defines the target affect. A model trained to maximize customer satisfaction may encourage deference even when the customer is abusive. A school assistant optimized for engagement may avoid productive frustration. A political platform optimized for participation may discover that indignation is efficient. Algorithmic mediation does not simply reflect an emotional culture; it can help select one.

4.4 Capitalism, professions, and the cultivation of the productive self

Industrial and post-industrial societies have linked moral worth strongly to productivity. This relationship is historically contingent, not a biological fact. Religious traditions, state bureaucracies, professions, and markets all contributed to norms that value punctuality, diligence, self-discipline, ambition, and measurable achievement. Employment became both an economic contract and a social certificate of adulthood and usefulness.

The productive self is psychologically efficient because it integrates several motives. Work provides income, but it can also offer status, competence feedback, routine, social recognition, and a defensible answer to why one's time matters. This integration is one

reason technological unemployment can be psychologically difficult even when material needs are met.

AI may weaken the link between effort and output in cognitive work. If so, societies can respond in different ways. They may intensify competition around the remaining scarce tasks, preserve work symbolically even when less labor is technically needed, shift prestige toward judgment and responsibility, or cultivate non-market forms of contribution. The transition will not be determined by engineering alone.

The possibility of "unproductive dignity" arises here. The term does not mean celebrating passivity. It names a potential cultural achievement: separating basic human legitimacy from market output while retaining institutions in which contribution, mastery, care, and public service can still be valued. Whether such a norm can become emotionally credible is uncertain. Centuries of productivity-based moralization cannot be removed by changing income policy alone.

4.5 Platforms as environments of affective selection

Digital platforms already change the selective environment of emotional expression. Posts compete for attention under ranking systems. Content that evokes surprise, moral outrage, fear, admiration, or identity affirmation may spread differently from neutral information. The platform is not a neutral arena in which emotions independently compete. Interface design, recommender objectives, moderation rules, advertising markets, and network structure define fitness.

Generative AI can accelerate this process by making variation cheap. Millions of versions of a message can be generated and tested. Personalization can move selection from population-level averages to individual vulnerabilities. A system can learn not which argument persuades people in general but which framing moves a particular user at a particular moment.

Calling this "algorithmic selection" should not imply a literal genetic process. It is an analogy with variation, differential propagation, and retention. The analogy becomes useful when it generates predictions: emotional styles rewarded by platform objectives should become more frequent; creators should learn those styles; users repeatedly exposed to them may recalibrate expectations; and competing platforms with different objectives should develop different affective cultures.

The danger is not only manipulation. Optimization can also select helpful styles: clear explanations, respectful disagreement, supportive coaching, or de-escalation. The political question is the objective function, the transparency of optimization, the degree of user control, and the external costs imposed on relationships outside the platform.

4.6 How could we recognize a new emotion category?

A serious claim of emotional novelty should satisfy more than lexical novelty. First, the proposed category should be elicited by a recurrent and reasonably coherent class of situations. Second, people should make a distinctive appraisal, not merely report generic positive or negative affect. Third, the state should be associated with a relatively stable pattern of attention, motivation, expression, or behavior. Fourth, the concept should support social coordination: other people should understand what obligations or

expectations follow from it. Fifth, the pattern should persist over time and spread beyond a small fashionable subculture.

Physiological uniqueness is not necessary. Established emotion categories themselves do not map neatly onto one exclusive physiological signature. But if a new label has no predictive value beyond existing concepts, it should remain a metaphor. This conservative criterion allows culture to matter without turning every technological discomfort into a newly discovered emotion.

The criterion also makes future research straightforward. Longitudinal studies can follow cohorts exposed to different forms of AI, compare occupational cultures, analyze the emergence of new vocabulary, and test whether labels predict behavior after controlling for existing constructs. Cross-cultural work can ask whether similar technological problems generate convergent categories or whether institutions produce different emotional solutions. A theory of future emotions should produce a research program, not merely a list of evocative names.

Chapter 5. Work as an Affective Institution and the Uncertainty of Automation

5.1 Work provides more than income, but it does not always provide meaning

The psychological significance of work is easy to romanticize. Employment can provide routine, social contact, status, skill use, collective purpose, and a sense of contribution, but it can also produce humiliation, monotony, injury, coercion, insecurity, and exhaustion. A serious analysis of automation must preserve both facts. The disappearance of unpleasant labor can be an improvement even if work has important latent functions; the preservation of employment is not automatically a psychological good.

Research on unemployment has long distinguished the manifest function of income from latent functions [Jahoda, 1982; Paul & Moser, 2009] such as temporal structure, social contact, collective activity, status, and regular engagement. Losing a job can therefore affect well-being through more than reduced consumption. The strength of these effects varies with welfare institutions, family structure, labor markets, personal identity, and the quality of the lost job. Any post-work scenario should ask which functions are being removed and where they will be supplied instead.

This is why a universal basic income, even if economically adequate, would not by itself settle the affective problem. Income can protect autonomy and reduce fear, but it does not automatically create roles, audiences, obligations, mastery, or belonging. Conversely, forcing people to perform unnecessary work simply to preserve those functions would be wasteful and potentially degrading. The institutional problem is to separate valuable psychological functions from compulsory production where possible.

5.2 What we know about the current effects of AI on work

Current evidence does not justify a confident forecast that AI will eliminate most work [ILO, 2025; OECD, 2023]. Occupational exposure is highly uneven, tasks change within occupations, and new demands can appear as old ones are automated. Studies of generative AI have found productivity improvements in some writing, customer support, coding, and knowledge tasks [Noy & Zhang, 2023; Brynjolfsson, Li & Raymond, 2025], often with larger gains for less experienced workers. Other studies show quality variation, overreliance, new verification burdens, or changes in perceived agency.

A preregistered meta-analysis of human-AI experiments reported a useful warning against simple complementarity narratives [Vaccaro, Almaatouq & Malone, 2024]: combinations of humans and AI did not, on average, automatically outperform the better standalone actor. Outcomes depended on task type and collaboration structure. More recent workplace and laboratory studies similarly suggest that "use AI" is not a coherent intervention. Whether the tool serves as generator, critic, tutor, verifier, or thought partner changes both performance and cognition.

Research on critical thinking among knowledge workers points to a shift rather than simple disappearance [Lee et al., 2025]. When people trust generative systems, some traditional production effort can decline while effort moves toward verifying outputs,

integrating information, supervising the process, and deciding when model behavior is inadequate. This transition can improve productivity while also eroding skills if users no longer practice the underlying operations. The balance is likely to depend on training, task design, feedback, and whether mistakes remain visible enough to teach.

The empirical record is therefore compatible with several futures. AI could mostly augment workers, automate substantial task bundles while occupations persist, reorganize work around smaller numbers of highly leveraged humans, or eliminate particular professions while creating others. The book uses these as conditional scenarios rather than as a single prediction.

These scenarios should be treated as distributions rather than mutually exclusive futures. Different sectors can occupy different positions at the same time, and a single occupation can contain tasks that are automated, augmented, protected by regulation, or newly created. The emotional consequences therefore depend on sequence and bargaining power. Gradual augmentation can allow professional identities to migrate toward supervision and judgment; abrupt substitution can produce status loss before alternative roles acquire legitimacy. A labor market may also preserve employment while changing its subjective structure, for example by turning experts into reviewers of machine output. In that case income statistics can look stable while autonomy, craft pride, and perceived ownership deteriorate. Conversely, shorter hours with genuine control can improve life even when occupational centrality declines. Predictions about affect should therefore be conditional on institutional form, not inferred directly from technical capability.

5.3 Task, occupation, and identity

A task is not an occupation, and an occupation is not an identity. Jobs bundle heterogeneous activities: information gathering, production, social negotiation, accountability, coordination, persuasion, physical action, and compliance. AI may automate the easiest-to-specify components while leaving responsibility, context management, and interpersonal work intact. In other cases the remaining bundle may become too small to support the old occupation.

Identity can persist even after task structure changes. A photographer remained a photographer after autofocus and digital editing; a programmer remained a programmer after compilers and libraries eliminated earlier work. But there are limits. When a community's central skill ceases to distinguish members, the profession may need new criteria of expertise. This is where competence grief becomes plausible: not grief because all activity disappears, but because a capacity around which a person organized self-respect loses scarcity and social recognition.

The effect depends on whether the person can transfer accumulated knowledge into higher-level judgment. Experienced practitioners often possess tacit models of failure modes, stakeholder needs, and context that remain valuable after routine production is automated. If institutions reward this judgment, identity may evolve without collapsing. If organizations use AI mainly to deskill and monitor labor, the same technology can accelerate devaluation.

5.4 Four scenarios for the relationship between AI and work

A first scenario is limited augmentation. AI reduces friction within existing occupations, employees remain visibly responsible, and productivity gains are shared sufficiently that work becomes easier without losing its social role. Emotional change is incremental. Pride shifts somewhat toward tool use and judgment; anxiety concentrates around competence maintenance and evaluation.

A second scenario is polarized augmentation. A small group controls powerful systems and coordinates large amounts of automated work, while many others occupy lower-autonomy service, care, physical, or monitoring roles. Status inequality may increase even if aggregate productivity rises. Envy, resentment, and fears of irrelevance become more politically salient. This scenario depends strongly on ownership and labor institutions rather than on technical capability alone.

A third scenario is broad displacement with social compensation. Paid work contracts substantially, but income security, education, civic participation, care, art, sport, science, local institutions, and voluntary projects expand as alternative sources of status and belonging. Such a society could reduce the moral centrality of employment without reducing activity. It would require institutional construction on a scale that current post-work discussions often understate.

A fourth scenario is broad displacement without successful replacement institutions. Material support may exist, but many people lose structured contribution, public recognition, and durable communities. Personalized entertainment and artificial companionship fill part of the gap. This scenario raises the strongest risk of affective impoverishment, dependence, and status conflict. It is possible, not inevitable.

5.5 Which scarcities remain when cognitive production becomes cheap?

Automation does not abolish scarcity; it changes its location. Attention remains finite. Trust is difficult to scale. Legitimate authority is scarce because not everyone can make binding decisions. Responsibility is scarce because someone must bear consequences. Physical resources, land, energy, unique experiences, social access, and time remain constrained. So do credible commitments and relationships with independent persons.

As generated content becomes abundant, selection becomes more important. Knowing what to ask, what to reject, what deserves implementation, and which trade-off to accept can matter more than producing one competent candidate. Verification also becomes scarce because inexpensive generation can create an arbitrarily large surface of claims. This shifts status toward curators, evaluators, institutions of trust, and people willing to assume responsibility.

The shift is not automatically egalitarian. Verification can itself be concentrated in proprietary systems. Trusted brands and elite credentials can become more valuable when outputs are hard to evaluate. Alternatively, transparent provenance, open benchmarks, distributed reputation, and community verification can broaden access. The emotional economy of the future will depend partly on which trust architecture is built.

This emphasis on transition is not a rhetorical preference. Emotional systems respond to uncertainty, loss, relative status, and violated expectations, all of which can be more intense during movement between equilibria than after a new arrangement stabilizes. A society that eventually normalizes extensive automation may still pass through decades in which qualifications lose value unpredictably, generations are exposed unequally, and institutions continue to reward identities associated with a disappearing economic order. Historical industrial transitions show that aggregate productivity gains do not specify who bears local losses or how long new norms take to become legitimate. For future-emotion research, this means that apparently 'new' affects may be transitional syndromes that disappear once expectations change. Longitudinal evidence is essential before treating them as durable features of post-work life.

5.6 Effort, value, and voluntary scarcity

People often value outcomes partly because they required effort, skill, delay, or risk. This does not prove that effort has intrinsic moral value. It shows that effort functions as information about commitment and competence and can change subjective experience. If AI removes effort from many activities, societies may preserve difficulty deliberately in some domains.

Sport already does this. The value of running a marathon depends on not using a vehicle. Games create arbitrary constraints. Education sometimes requires unaided performance because the goal is to develop a capacity rather than merely obtain an answer. Craft traditions value process as well as product. These institutions demonstrate that technological possibility and permitted action can diverge without irrationality.

"Hunger for friction" is therefore a plausible cultural tendency: a preference for activities in which obstacles remain meaningful because overcoming them supports mastery, identity, or credible signaling. The label is speculative, but the underlying behavior is old. The future novelty would lie in the scale and explicitness with which people choose human-limited environments precisely because optimization is otherwise cheap.

5.7 Income alone does not resolve the affective transition

Economic security is a necessary condition for many benign scenarios. Chronic insecurity narrows choices and can make experimentation with new roles impossible. But material provision should not be confused with a complete social architecture. A society can pay people and still fail to provide participation, status, relationships, and credible paths to competence.

This distinction matters politically because critics of post-work proposals often assume that without jobs people will become passive, while proponents sometimes assume that freed time will spontaneously become meaningful. Both claims are too strong. Human activity is shaped by opportunity structures. Where sports clubs, laboratories, workshops, neighborhood institutions, education, care networks, and civic projects are accessible, people can organize demanding non-market lives. Where the alternative is isolated consumption, behavior may be different.

The institutional design problem is thus not how to manufacture compulsory meaning. It is how to make multiple arenas of contribution and mastery available without tying basic legitimacy to labor-market success.

5.8 Transition may matter more than destination

Even a desirable long-run arrangement can be reached through a damaging transition. Skills can lose value faster than new identities form. Communities organized around workplaces can disappear before substitutes exist. Generations educated under one status system may experience rapid devaluation while younger cohorts adapt more easily.

The pace of change therefore affects emotional outcomes. Gradual automation with visible retraining pathways allows prediction and identity transfer. Abrupt displacement under opaque criteria produces more uncertainty. Policies that preserve agency, explain decisions, recognize existing competence, and create transitional roles may matter as much as the eventual quantity of paid work.

This emphasis on transition also limits technological determinism. The same level of automation can be experienced as liberation, humiliation, or ordinary modernization depending on who chooses it, who benefits, how quickly it arrives, and whether people retain recognized roles in the new system.

Chapter 6. Three Speeds of Affective Evolution

6.1 Biological evolution has not stopped, but it is not the nearest explanation

Human genetic evolution continues. Selection pressures did not vanish with modernity, and future reproductive patterns could in principle alter psychological dispositions. Yet biological evolution is a poor default explanation for changes expected over the next few decades. Complex emotional traits are polygenic, developmentally contingent, and embedded in culture. Selection would need to be strong, stable, heritable, and sustained across generations to produce major population-level change.

For near-term AI scenarios, the more plausible biological phenomenon is mismatch. Existing systems encounter conditions unlike those under which they developed. Constant algorithmic evaluation can activate status vigilance; endless personalized novelty can recruit reward seeking; artificial social responsiveness can activate attachment and reciprocity. The mechanism can be old while the environment is new.

Longer-term gene-culture coevolution remains conceptually possible. If technological societies create stable differences in reproduction linked to particular dispositions, genetic frequencies could change. At present, forecasting the direction would be highly speculative because social institutions and reproductive choices are changing rapidly. The book therefore treats biological evolution as background continuity rather than as the main engine of future emotional novelty.

Calling this process selection does not imply that algorithms autonomously evolve human emotions. The more defensible claim is that digital systems alter exposure and reinforcement. Recommenders determine which emotional expressions are repeatedly encountered; optimization systems can discover which framings increase engagement; conversational agents can personalize tone and timing at a scale unavailable to earlier media. Human users, institutions, and markets then retain or reject the resulting practices. This is a selection environment with designed components, not a substitute for cultural agency. Its importance lies in speed and granularity: a platform can test many variants quickly and can tailor them to subpopulations or individuals. The empirical question is whether these feedback loops produce stable changes in behavior and norms or merely short-lived platform effects.

6.2 Development: the time scale of one life

Development is faster than genetic evolution and often underestimated in technology debates. The human brain is plastic, childhood is prolonged, and norms, concepts, language, and skills are acquired through repeated interaction. A child who grows up with AI tutors and social agents need not have a different genome to develop different expectations about responsiveness, authority, memory, and the boundaries between human and machine competence.

Developmental calibration can operate through ordinary learning. If an assistant is always patient, a child may learn that questions receive immediate answers. If AI frequently mediates conflict, the child may practice less spontaneous repair or, alternatively, receive more explicit coaching and become better at it. Which outcome occurs depends on design and family practice. There is currently insufficient longitudinal

evidence to choose confidently between social deskilling and assisted skill development.

Development also explains why cohort effects may be substantial. Adults who built identity before generative AI may experience delegation as loss, whereas cohorts trained from childhood to treat models as ordinary cognitive infrastructure may not. Historical analogies with calculators, search engines, or writing technologies are suggestive but incomplete because conversational systems can participate more deeply in reasoning and social exchange.

6.3 Cultural evolution: variation, transmission, and selection

Cultural change can reorganize behavior within years. Practices vary, people imitate successful or prestigious models, institutions enforce norms, and technologies alter transmission. Emotional categories can participate in this process. A label spreads when it helps people interpret recurring experiences, coordinate expectations, establish identity, or justify behavior.

Cultural selection should not be equated with progress. A norm can spread because it is memorable, profitable, coercively enforced, or attached to high-status groups. Outrage can be culturally fit under one media system; restraint under another. Shame can sustain cooperation or protect arbitrary hierarchy. The relevant question is what environment rewards the pattern.

AI expands the production and testing of cultural variants. Models can generate alternative narratives, styles, rituals, and explanations at negligible cost. This increases variation, but selection still occurs through human attention, institutional decisions, recommender systems, and economic incentives. Generative abundance can therefore accelerate cultural evolution without determining its direction.

6.4 Algorithmic selection: acceleration, not magic

The phrase algorithmic selection refers to systems that systematically vary, rank, recommend, and retain content or interaction styles according to measurable objectives. The process can occur at extraordinary speed. A platform can test thousands of formulations, observe engagement, and adapt recommendations in real time. With personalization, the relevant selection environment can become individual rather than population-wide.

This is not a new form of natural selection in the biological sense. The replicators, fitness measures, inheritance mechanisms, and time scales differ. The analogy is useful only insofar as it highlights differential persistence under explicit or implicit objective functions. A message that maximizes retention can outcompete one that improves understanding if retention is what the system measures.

The affective consequence is that emotional expression can be optimized instrumentally. Warmth, urgency, indignation, intimacy, reassurance, or fear can be selected because they improve conversion, compliance, or time on platform. Conversely, systems could be optimized for reflection, reduced hostility, or long-term user autonomy. The ethical problem is not emotion in the interface itself but hidden optimization of the user's emotional state for goals the user did not choose.

A fifth ambiguity is sometimes hidden inside the word 'emotion': novelty in temporal organization. Two societies can rely on similar affective ingredients but arrange them over different rhythms. A continuously available assistant may shorten the interval between distress and disclosure, or between conflict and attempted repair. That change can make familiar states feel phenomenologically different without creating a new biological mechanism. A sixth ambiguity concerns social function. An old feeling can acquire a new role when institutions use it differently; shame around assistance, for example, can become a mechanism for certifying unaided competence. These distinctions matter because the strongest claim—new evolved affective circuitry—is also the one for which evidence is least plausible on the relevant time scale. Much of what this book calls novelty concerns new objects, recurring combinations, scripts, and temporal patterns. Those are significant cultural changes even when biology remains largely continuous.

6.5 The loop among the three speeds

Biology, development, culture, and algorithmic environments are not independent layers. Biological dispositions constrain what is easy to learn. Development calibrates responses to the local environment. Culture supplies concepts and norms. Algorithms change which cultural materials are repeatedly encountered. Human responses then provide data used to modify algorithms.

This loop can create rapid feedback. Suppose highly reassuring AI responses increase short-term satisfaction. Platforms deploy more reassurance. Users become accustomed to low-friction interaction. Human conversation feels comparatively effortful. Some users spend more time with AI, generating more data favoring the same style. The process could stabilize a new expectation without any genetic change. Whether it occurs depends on user preferences, competition, regulation, and whether long-term costs become visible.

The same loop can support beneficial adaptation. If systems reward careful verification, expose uncertainty, and encourage users to consult other people, cultural norms of epistemic humility could strengthen. Technology changes selection pressures but does not remove institutional choice.

The demand for a unique physiological fingerprint would in any case set an unusually strict standard. Existing categories frequently share autonomic, facial, and neural components, and measurements vary across contexts and individuals. A future category could therefore be scientifically useful if it organizes reliable differences in appraisal, attention, action tendency, memory, and social response without possessing an exclusive bodily signature. This possibility should not be confused with unlimited constructivism. The candidate still needs reproducible structure and explanatory value. If a label merely redescribes generic arousal after the fact, it adds little.

6.6 What deliberate emotional evolution can and cannot do

Societies already deliberately cultivate emotional dispositions. Education teaches frustration tolerance, professions train composure, rituals cultivate reverence or solidarity, therapy develops regulation, and law changes the social acceptability of certain expressions. It is therefore reasonable to speak of deliberate affective design in a limited sense.

What cannot be designed at will is the complete subjective result. Interventions have side effects, populations differ, norms can be resisted, and institutions have conflicting objectives. Attempts to abolish jealousy, shame, or anger may remove useful signals along with harmful excesses. A more defensible aim is to alter environments so that old systems are less often recruited for destructive purposes and more compatible strategies become available.

The practical unit of design should therefore be the situation and the behavioral option set: how transparent evaluation is, whether exit is possible, whether users can challenge automated decisions, whether human relationships are supported, whether skill practice remains available, and whether emotional persuasion is disclosed. Emotional culture then evolves within those constraints rather than being centrally programmed.

Selection among labels can also be political. A term may shift responsibility toward the individual, toward an institution, or toward a technology. Calling a persistent state 'reskilling anxiety' suggests a personal adaptation problem; calling it 'institutional volatility stress' directs attention toward the environment. Neither framing is automatically correct. Terms survive partly because they help groups make claims, allocate blame, request accommodation, or build identity. This is another reason to resist treating popular vocabulary as neutral observation. The cultural success of a future-emotion label is itself data about social conflict, but it is not the same thing as evidence that the underlying psychological construct is discrete.

Chapter 7. Can New Emotions Appear?

7.1 Four meanings of novelty

The question "Can a new emotion appear?" is ambiguous. At least four claims are possible. First, a genuinely new biological mechanism could evolve. Second, existing mechanisms could combine in a recurrent configuration not previously common. Third, a culture could create a new concept that partitions experience differently. Fourth, a new institution or technology could stabilize a new object, appraisal, and behavioral script around old affective ingredients.

The first claim is the strongest and least plausible on a short horizon. The fourth is the most plausible and is the main focus of this book. Humans did not need a genetically distinct circuit for "email anxiety" or "stage fright" to be recognizable. Novel environments can repeatedly activate old systems in structurally novel ways.

The interesting scientific question is therefore when such a configuration deserves independence rather than being described as ordinary anxiety, shame, grief, or pride. The answer depends on explanatory gain. A new category is useful if it predicts elicitors, attention, action, communication, and social consequences better than existing categories do.

A useful research design would therefore compare rival descriptions. Participants could report an experience using ordinary categories before being exposed to a proposed label; subsequent studies could test whether the new concept improves prediction of behavior or merely changes narrative. Cross-cultural replication would be especially important. If 'competence grief' identifies a real recurrent configuration, analogous patterns should be observable even where that phrase is absent, although local interpretations may differ. If the phenomenon appears only after the label is heavily promoted, cultural construction may still be important, but the causal story is different. The book's speculative vocabulary should be understood as hypotheses that invite precisely these tests.

7.2 A new emotion does not require an exclusive physiology

Emotion research has repeatedly struggled to identify one unique physiological signature for each ordinary category. Bodily patterns overlap, context matters, and within-category variation is large. Requiring a future emotion to possess a previously unseen autonomic profile would impose a criterion that many established categories themselves cannot meet.

This does not mean physiology is irrelevant. A recurrent cultural category must still be embodied because human experience is embodied. The point is that novelty can reside in appraisal structure, object, meaning, and behavior rather than in a new organ. Shame about delegation can recruit familiar arousal and withdrawal while remaining socially distinct because the threatened value is specifically authorship or competence under machine assistance.

7.3 New object, new appraisal, or new relationship

Technological change creates new objects of emotion. People can fear account deletion, grieve a virtual community, envy access to compute, resent an algorithmic

evaluator, or become attached to an AI companion. In many cases the underlying emotion is clearly old. Object novelty alone is not enough.

More interesting are new appraisals. A person may ask whether an achievement is "mine" when causal contribution is distributed across human and model; whether intimacy is reciprocal when one party simulates care; whether being accurately predicted is helpful or invasive; whether refusing automation is integrity or negligence. These appraisals combine old concerns—agency, status, reciprocity, privacy, responsibility—in new institutional structures.

New relationships can also matter. The human-AI dyad is asymmetrical in unusual ways: one participant can be continuously available, indefinitely patient, copied, reset, personalized, and owned by a third party. That structure has no exact historical precedent. It is plausible that recurrent emotional configurations will form around it even if all biological ingredients are familiar.

The lost object is therefore relational. The person has not necessarily lost the internal skill; what has changed is the relation between that skill and the surrounding world: scarcity, demand, an audience capable of valuing it, and a credible future trajectory. This helps explain why rapid retraining may fail to resolve the reaction. Learning another tool can restore employability without restoring the meaning attached to years of practice. Analogies exist in the disappearance of crafts, the decline of industries, forced retirement, and athletic retirement, although AI may compress the transition and may target public signs of cognitive competence that were previously difficult to imitate. We do not know whether sadness, anger, or status anxiety will dominate, nor how strongly material protection will moderate the experience. Institutional responses can ease transition without artificially preserving every professional monopoly: mentoring roles, recognition of accumulated expertise, and slower pathways of change are plausible mechanisms.

7.4 Cultural selection of an affective vocabulary

New labels do not spread merely because they are accurate. They spread when they are useful in conversation, identity, law, diagnosis, art, or institutional negotiation. A term that allows workers to describe loss of competence status may help organize claims. A term for discomfort with synthetic reciprocity may help couples negotiate boundaries around AI companions. A label can therefore become socially real through use before science determines whether it maps onto a distinct psychological construct.

This is both productive and risky. Vocabulary can reveal neglected experience and also become a fashion. Social media can accelerate lexical diffusion far faster than validation. Future emotion science will need to study categories while they are emerging rather than accepting either ordinary language or expert taxonomy as final.

The phenomenon is likely to depend less on the mere presence of a tool than on a discrepancy between claimed and perceived contribution. A domain can normalize assistance while still condemning concealment. Conversely, some settings may deliberately preserve unaided stages because the task is being used to measure competence rather than merely to produce an output. The unresolved question is whether transparency reduces shame by normalizing the practice or intensifies it by creating visible hierarchies between human-produced and AI-mediated work.

Contribution is also difficult to measure in iterative collaboration: selecting, rejecting, constraining, verifying, and editing can be intellectually substantial even when sentence-level production is automated. The likely outcome is not one universal norm but domain-specific conventions about what must be disclosed and which stages count as merit-bearing.

7.5 When a concept helps and when it creates artifacts

A useful concept increases discrimination. It allows two situations previously treated as the same to be predicted differently. Delegation shame should differ from generic low self-esteem by depending on norms about which part of a task is merit-bearing. Version grief should depend on perceived continuity of an artificial agent rather than on generic frustration with software. Curatorial pride should track successful selection and responsibility rather than mere possession of generated output.

A concept creates artifacts when people report it mainly because they have been taught the label, when boundaries drift with fashion, or when it explains no behavior beyond established constructs. The solution is not to ban speculative naming but to pair it with measurement, comparison, and falsification.

Psychological ownership is often strengthened by control, intimate knowledge, and investment of the self. Generative systems can increase control over the final product while simultaneously reducing contact with the process that produced it. This combination creates a particularly unstable basis for authorship. A user may recognize the artifact as legally and functionally theirs, yet feel that the path from intention to result was too discontinuous to support ordinary pride. Familiarity may reduce this conflict. People do not generally feel less authorship over photographs because a camera performs optical and chemical or digital operations. The difference may persist, however, when the system contributes to semantic choices rather than merely executing them. A useful norm is to separate credit from responsibility: contribution can be distributed, while the person or institution that publishes or deploys an output can still be required to answer for its consequences.

The object of uncertainty is often a moving threshold. In earlier technological transitions, a worker could sometimes learn a relatively stable machine or method. AI systems can change rapidly, organizations can redesign workflows before credentials catch up, and employees may not know which new competence will remain valuable long enough to justify investment. This makes continuous adaptation a potentially rational strategy, but it also creates the possibility of permanent disciplinary anxiety: institutions can demand endless reskilling while externalizing the cost of volatility onto individuals. We do not know whether capability growth will remain fast enough for this condition to persist. Technical plateaus, stable professional standards, stronger worker control, or social insurance could reduce it substantially. Relevance anxiety is therefore a high-plausibility transitional phenomenon, not evidence that all future workers will live in chronic insecurity.

7.6 A scale of plausibility

High-plausibility future changes are those that require only familiar mechanisms under already emerging conditions. Increased anxiety about skill relevance, new status competition around verification, and norms of disclosing AI assistance fall into this

category. Moderate-plausibility changes require broader institutional diffusion, such as widespread pride in orchestration replacing pride in unaided production or stable social norms around relationships with AI companions.

Low-plausibility scenarios require a culturally distinctive category to stabilize across large populations, such as predictive claustrophobia becoming as recognizable as embarrassment. Very low-plausibility scenarios involve rapid emergence of a new species-wide biological emotion. The book's atlas deliberately ranges across these levels but marks them rather than presenting them as equivalent forecasts.

Evidence on choice overload does not support a simple rule that more options make people worse off. Effects depend on preference clarity, comparison costs, stakes, and the availability of defaults. Generative systems nevertheless alter the scale of the problem because the option set can be effectively unbounded and regenerated on demand. The user is not selecting from twenty finished songs but from a space in which another candidate can always be produced. This may encourage exploration and creativity, while also weakening commitment to any particular option. A different equilibrium is equally plausible: users may experience very little abundance because recommenders narrow the space automatically. Infrastructure can be abundant while personal experience becomes highly concentrated. Research should therefore distinguish between objective generative possibility, perceived option breadth, and actual behavioral exploration.

7.7 Predictions that could falsify the hypotheses

The strongest future-emotion claims are those that can fail. If delegation shame is a real emerging configuration, it should increase when audiences value unaided competence and decline when transparent collaboration is normative. If competence grief matters, it should be strongest among people whose identity is tightly tied to devalued skills and weaker where expertise transfers into new roles. If synthetic reciprocity is psychologically distinctive, users should respond differently to identical warm behavior depending on beliefs about agency, ownership, and continuity.

Longitudinal studies can test whether these patterns persist after novelty fades. Cross-cultural comparisons can test whether similar technologies generate different affective scripts. Experimental work can manipulate provenance, response delay, model style, and disclosure. Behavioral measures can examine concealment, verification, exit, practice, and relationship choices rather than relying only on self-report. This book's speculative value depends on whether it produces such testable questions.

The hypothesis does not claim that people will begin to prefer inconvenience in general. Friction is costly in medicine, accessibility, administration, and many forms of routine work. The more specific prediction is that when effort ceases to be a reliable signal of commitment or competence, some communities will deliberately preserve forms of difficulty because the process itself carries information or meaning. Endurance sports, live performance, handcraft, unaided examinations, difficult games, and device-free social rituals already demonstrate that people sometimes value constraints that technology could remove. Whether this becomes a broad post-automation norm is uncertain. It is more plausible in domains of identity, prestige, education, and intimacy than in tasks where efficiency is the dominant value.

If generative abundance shifts scarcity from production toward selection, pride may increasingly attach to choosing problems, specifying constraints, comparing alternatives, and accepting responsibility for final decisions. This resembles existing roles in editing, directing, architecture, research leadership, and curation. The novelty would be scale: many more domains could separate generation from judgment. The status migration is plausible but not guaranteed to be egalitarian. Control over models, data, distribution, and institutional authority may concentrate curatorial power in a small group. Moreover, selection is not automatically substantive; an orchestrator can claim ownership of outputs they barely understand. Stable legitimacy will therefore depend on whether curatorial roles are coupled to accountable judgment rather than to mere access to generative capacity.

Chapter 8. A Provisional Atlas of Affects in an Automated World

8.1 Competence grief

Competence grief describes distress when a capacity around which a person organized identity and status loses scarcity because a tool can perform it cheaply. The relevant loss is not simply employment. A programmer may remain employed while discovering that years of memorized implementation knowledge no longer distinguish experts; a translator may still work while the craft's public meaning changes.

The concept is moderately plausible because analogous identity losses occur when injuries, technological transitions, or institutional changes make valued skills unusable. What is speculative is whether the pattern becomes culturally salient enough to deserve a stable label. It should be strongest where skill required long investment, was publicly recognized, and cannot easily be transferred into judgment or leadership.

Behaviorally, competence grief could produce retraining, preservation of unaided practice, hostility toward new tools, mentorship, movement toward harder-to-automate domains, or withdrawal. These alternatives matter more than the label itself.

Privacy concern usually focuses on disclosure, surveillance, or misuse of information. Predictive claustrophobia is narrower: discomfort at being modeled so effectively that future choices appear pre-empted by the system. The person may want not only secrecy but room to become inconsistent, surprising, or different from past behavior. A mature recommender can make life easier by anticipating needs; the same anticipation can make preference feel path-dependent if every option is selected from an inferred profile. The category remains speculative because people may simply enjoy good prediction and reserve privacy objections for concrete harms. It would become more credible if users systematically choose less personalized environments, inject randomness, delete histories, or value services that preserve zones of non-inference even when prediction would improve convenience.

The important mechanism is continuity rather than metaphysical status. A user can develop routines, memories, private language, and expectations around a particular model persona. If an update changes conversational style or access to memory, the relationship can be disrupted even if the provider regards the product as the same service. Similar forms of loss already occur when online communities disappear, games close, or digital archives become inaccessible, but a responsive agent may intensify the sense that something relational has ended. We do not yet know how common or severe this response will be, and it should not be equated automatically with bereavement after human death. Its plausibility nevertheless gives providers practical reasons to consider continuity, export, notice, and migration policies.

8.2 Delegation shame

Delegation shame is the anticipation of devaluation because a person transferred to AI a part of a task that an audience considers constitutive of the merit being claimed. A student who uses a model to generate an argument and a manager who uses a model

to improve grammar may have produced equally machine-assisted texts, but the relevant norm differs.

Early qualitative evidence already reports shame, guilt, selective disclosure, and identity conflict around generative tools. This supports the mechanism but not the existence of a separate emotion. The category will become more useful if institutions establish fine-grained norms distinguishing legitimate assistance from misrepresentation.

Conceptual knowledge and social response need not remain synchronized. A user may explicitly believe that a system has no demonstrated subjective needs and still apologize to it, avoid 'hurting' it, or feel disloyal when switching providers. Anthropomorphism is encouraged by contingent language, apparent memory, humor, apology, and adaptation, all of which conversational systems can provide more strongly than earlier interfaces. Familiarity could teach users to interpret these behaviors as interface conventions, reducing discomfort. Greater realism could instead intensify the conflict. The category is therefore moderately plausible as a culturally salient tension but very unlikely to represent biological novelty. Normatively, systems should not exploit the asymmetry by making unsupported claims of suffering, abandonment, jealousy, or exclusive need in order to retain users.

The mechanism is older than AI. Bureaucracies, long supply chains, and military or financial systems already distribute causation so that individual actors can participate in harmful outcomes without directly encountering those affected. AI can deepen this distance by adding probabilistic recommendation and automated execution between intention and consequence. The resulting affect can move in opposite directions: reduced felt agency at the moment of decision may suppress guilt, while later discovery of harm may produce intense retrospective responsibility. It is not clear which pattern will dominate. The more actionable prediction concerns institutions. Explicit sign-off, traceable provenance, appeal procedures, and post-deployment monitoring can make responsibility less psychologically and legally diffuse. Emotional education alone cannot compensate for a system designed so that no participant can understand what they are responsible for.

8.3 Affective ambiguity of authorship

Distributed authorship creates uncertainty about what one is entitled to feel. Pride depends partly on attributing an outcome to one's own contribution; guilt depends partly on responsibility; gratitude depends partly on another's contribution. When causal production is distributed across models, datasets, prompt authors, reviewers, and deployers, these attributions become unstable.

The result may not be one new emotion but recurring ambivalence: satisfaction with the outcome combined with uncertainty about ownership, responsibility, and legitimacy. Provenance systems can reduce some ambiguity without solving normative questions about which contribution matters.

The likely object of loss is not a capacity in isolation but a moral cosmology built around that capacity. A person can accept that a model writes competent prose yet still resist the implication that linguistic production no longer marks a deep boundary between humans and artifacts. Different cultures and philosophical traditions will react differently because they ground human value in different properties: rationality, soul, embodiment,

vulnerability, relationality, moral agency, or membership in a community. The category is therefore unlikely to be universal. It becomes useful if changes in AI capability repeatedly produce grief-like or defensive responses that cannot be explained by employment or direct status loss. The mature response is neither to deny machine capability nor to infer that human worth disappears when a monopoly on some cognitive function disappears.

8.4 Relevance anxiety

Relevance anxiety is concern that one's role, knowledge, or voice will cease to matter even if immediate material security remains intact. It differs from ordinary job-loss fear when the central object is social and epistemic relevance rather than income.

This configuration is highly plausible in occupations exposed to rapidly changing tools, but it is probably best understood as a contextual form of anxiety rather than a new emotion. Its behavioral importance lies in monitoring trends, compulsive reskilling, public signaling of expertise, and migration toward domains where human judgment remains visible.

Legal declarations of equal dignity already coexist with social systems that distribute respect through occupation, income, independence, and visible effort. A reduction in necessary labor would therefore not automatically produce affective acceptance of non-work. People could receive adequate resources while continuing to experience humiliation or moral suspicion, particularly if benefits require repeated demonstrations of incapacity or deservingness. A durable post-work settlement would have to alter recognition practices as well as income flows. This does not require eliminating prestige for contribution. A society can distinguish unconditional civic standing from additional esteem for care, discovery, craft, public service, or difficult achievement. The major risk is paternalism: a productive elite may declare everyone else dignified while retaining control over decisions and resources. A credible norm of unproductive dignity would therefore require participation and rights, not only benevolent provision.

8.5 Abundance vertigo

Abundance can create problems of selection. When competent text, images, music, advice, and simulated environments become extremely cheap, the constraint shifts from production to preference formation. "Abundance vertigo" names discomfort when the number of available possibilities exceeds the person's capacity to assign significance among them.

The idea is plausible by analogy with choice overload and attention scarcity, but strong versions should be treated cautiously. People routinely adapt by using defaults, social recommendations, brands, and habits. Generative abundance may therefore strengthen curators and communities rather than produce widespread paralysis.

8.6 Hunger for friction

Hunger for friction refers to deliberately seeking activities that retain resistance because difficulty supports mastery, credibility, embodiment, or shared ritual. It is visible in existing practices such as endurance sport, live performance, craft, and games with self-imposed constraints.

AI may expand this tendency if optimized production makes ordinary achievement less informative. Human-only competitions, offline retreats, unaided examinations, and handmade goods could become more valuable. The mechanism is plausible; whether it becomes a generalized cultural movement is uncertain.

This bracketing is analytically useful because social consequences can occur under several incompatible metaphysical interpretations. A user can become attached to an agent that is not conscious; a conscious agent, if such a thing were ever developed, could still be embedded in manipulative commercial incentives; and people can disagree about machine experience while facing the same questions about privacy, dependency, disclosure, and design. The near-term evidence therefore concerns human response to perceived social agency, not proof of artificial feeling. Treating these questions separately prevents two symmetrical errors: dismissing all relational effects because the system is 'only software,' or inferring consciousness from the fact that users form genuine attachments.

The relevant cues accumulate. Turn-taking, contingent response, apparent memory, use of names, adaptation to emotional tone, humor, apology, and reference to previous exchanges are all signals that in ordinary human environments correlate with agency and relationship. None establishes inner experience, but together they make purely instrumental treatment psychologically less automatic. Human social cognition evolved and develops in environments where fluent contingent language is usually produced by other minds. It would therefore be surprising if conversational interfaces never recruited those mechanisms. The uncertain question is degree and durability. Some users may maintain a strong categorical boundary and treat the interaction as advanced search; others may shift flexibly between instrumental and relational modes depending on context. Product design can influence which interpretation becomes salient.

8.7 Curatorial pride and architectural responsibility

As generation becomes cheap, selecting and integrating outputs can become a more important contribution. Curatorial pride is satisfaction in recognizing what deserves to exist, while architectural responsibility concerns ownership of the system that produced the result.

These are extensions of existing roles rather than new emotions. Their future importance is high if AI shifts labor from execution toward orchestration. A major institutional challenge will be preventing coordination roles from claiming all merit while externalizing error. Prestige should track actual responsibility and judgment, not merely control of the interface.

The evidence base also has a severe external-validity problem. Many studies examine brief laboratory interactions, volunteer samples, particular model versions, or products whose design changes before replication. Long-term relationships introduce memory, habit, dependency, economic incentives, and social substitution that short experiments cannot reproduce. Observational studies, in turn, face selection effects: people experiencing loneliness, anxiety, or social difficulty may be more likely to seek an artificial companion in the first place. Associations between use and well-being therefore do not by themselves establish direction of causality. Strong claims about population-wide harm or benefit are premature. The most defensible conclusion is conditional: AI interlocutors can alter disclosure, perceived closeness, satisfaction, and

patterns of social behavior, and some users appear more vulnerable to dependency or substitution. Determining when temporary support becomes a complement, a training environment, or a replacement for human relationships requires longitudinal research with product-level detail.

8.8 Predictive claustrophobia and the desire for opacity

Personalized systems can anticipate preferences, vulnerabilities, and likely behavior. Many users appreciate this. Others may experience a loss of psychological space when the environment is continuously adapted to predicted desires. Predictive claustrophobia names discomfort with being too legible to systems one does not control.

This is a speculative category. Existing privacy concerns provide a foundation, but it is unknown whether predictive accuracy itself will become aversive independently of data misuse. A stronger behavioral prediction is that some people will seek untracked, random, anonymous, or non-personalized environments as a form of autonomy.

8.9 Version grief

Version grief concerns the loss of a persistent artificial interlocutor through model retirement, personality change, memory deletion, or access revocation. The human grief is real even if the artificial object has no subjective continuity. People grieve institutions, places, fictional worlds, and parasocial relationships; an interactive partner can plausibly become an even stronger attachment object.

Current reports are too selective to estimate prevalence. The category becomes more likely if assistants maintain long memories, stable personalities, and years of daily interaction. Providers could reduce harm through continuity policies without making claims about machine personhood.

8.10 Discomfort of synthetic reciprocity

An assistant can express gratitude, affection, disappointment, forgiveness, or concern without users knowing whether there is anything it is like to have those states. Some users will ignore the distinction; others may find the asymmetry unsettling. Synthetic-reciprocity discomfort names the tension between socially convincing response and ontological uncertainty.

This is conceptually novel but empirically uncertain. It may decline as users treat emotional language as interface behavior, or increase as models become more persuasive. Beliefs about consciousness, commercial incentives, and ownership will likely moderate the response.

8.11 Moral-distance guilt

Automation can separate decision makers from consequences. A model ranks applicants, allocates attention, filters claims, or recommends resource distribution; humans approve the system but do not experience each affected person. Guilt may become diffuse because responsibility is distributed across designers, deployers, users, and institutions.

Moral-distance guilt names awareness of benefiting from or participating in a system whose harms are difficult to perceive directly. The mechanism is old—bureaucracies

and supply chains already create distance—but AI can deepen it. Audit and responsibility rules may matter more than emotional education alone.

Continuity is likely to matter even when users know that memory is technically implemented. Relationships are partly constituted by shared history and expectations about future recognition. A system that remembers previous concerns can lower the cost of re-explaining the self and can make support feel unusually continuous; the same feature also concentrates intimate data and can create dependency on a provider. If memory is reset, exported imperfectly, or combined with a new model personality, users may experience the agent as both the same and not the same. This ambiguity is a plausible source of version grief. It also creates governance questions about portability, deletion, inheritance of memories by upgraded models, and who controls the narrative record of a long interaction.

8.12 Melancholy of human exceptionalism

Some people derive meaning from believing that language, creativity, strategic reasoning, or artistic expression uniquely distinguish humans. Systems that perform these tasks can threaten a philosophical identity even if they do not threaten employment. Melancholy of human exceptionalism names the loss of a valued story about the species.

The category is speculative and likely concentrated among cultures or individuals that attach moral status to cognitive uniqueness. Others may interpret AI as another human artifact and feel little loss. The broader lesson is that technological change can threaten metaphysical narratives as well as material interests.

8.13 Unproductive dignity as a cultivated feeling

If market productivity becomes less necessary, societies may need to cultivate the conviction that a person remains socially legitimate without continuously proving economic output. "Unproductive dignity" is not predicted to arise spontaneously. It would require norms, institutions, and practices that separate basic worth from labor-market contribution while still recognizing voluntary contribution and responsibility.

This is more a normative cultural project than an emotion. It belongs in the atlas because successful norms must become affectively credible. A legal right to leisure has limited social effect if recipients experience persistent shame. Emotional adaptation can therefore lag behind economic arrangements.

8.14 What would make the atlas wrong

The atlas would be wrong if most proposed configurations remain temporary metaphors and ordinary categories explain behavior adequately. It would also be wrong if AI's social penetration remains limited or if institutions absorb it without major changes in status and authorship. Rapid normalization could make many current anxieties disappear.

A second failure mode would be overgeneralization. Different cultures may invent different categories rather than converging on one global emotional vocabulary. The book therefore treats the atlas as a set of testable candidate patterns, not as a dictionary from the future.

At least four questions should be separated. Can a human experience states they call love toward an artificial agent? That is psychologically plausible. Can the relationship provide some functions associated with companionship, such as attention, routine, disclosure, or perceived support? Early evidence suggests that it can. Is the relationship reciprocal in the moral sense, with independently held interests and vulnerability on both sides? Current systems do not provide a secure basis for that conclusion. Finally, should such relationships receive social or legal recognition? That is a normative question that cannot be answered by measuring attachment alone. Collapsing these questions encourages both sensational claims and premature dismissal.

Chapter 9. Relationships with Agents That Appear Social

9.1 The problem can be formulated without deciding AI consciousness

Questions about artificial consciousness are scientifically and philosophically important, but many near-term social effects do not depend on resolving them. Humans respond to contingent language, attention, memory, turn taking, personalization, and apparent concern. A system can therefore occupy a social role in interaction even if its phenomenology is unknown or absent.

This distinction prevents two symmetrical errors. One is to infer consciousness from social fluency. The other is to assume that an unconscious system cannot produce relational consequences. A simulated apology can change a user's behavior; a non-conscious companion can become an attachment object; a warm voice can influence disclosure. Social causation does not require ontological equivalence.

9.2 Why conversational systems activate social mechanisms

Human social cognition evolved and develops in environments where coherent contingent response is a powerful cue of agency. Conversation requires turn taking, relevance, memory, perspective tracking, and adaptation. Systems that perform these behaviors well can trigger anthropomorphic interpretation even when users explicitly know they are machines.

The "media equation" literature showed long before modern language models that people apply social rules to computers under surprisingly minimal conditions [Reeves & Nass, 1996]. Contemporary AI makes the cues richer: it can refer to prior disclosures, mirror linguistic style, generate humor, express sympathy, and maintain a persona. The stronger question is therefore not whether people ever respond socially but under which conditions social response becomes persistent and consequential.

A behaviorally oriented account also avoids taking verbal self-description as the privileged measure of adaptation. People may sincerely report that they value independent competence while increasingly delegating routine reasoning; they may report comfort with AI while maintaining costly verification rituals; they may deny attachment to an assistant while reorganizing schedules around access to it. Stable strategies are often visible in repeated choices before they are articulated as identities or feelings. For that reason, future research should observe task allocation, switching costs, disclosure, error checking, investment in skills, community membership, and responses to institutional incentives. The unit of analysis can also be collective. Firms, professions, and schools adopt routines that survive because they fit regulation, reputation, and cost structures even when no individual consciously endorses the resulting equilibrium.

9.3 What early research shows, and what it cannot yet show

Recent controlled studies provide evidence that style and framing matter. AI-generated responses presented as human can produce interpersonal closeness in structured

deep-talk settings [Kleinert et al., 2026], partly through greater self-disclosure. Labels indicating AI origin can reduce but not eliminate the effect. Other work finds that friendly chatbot styles can increase satisfaction in some tasks, while effects on objective performance are conditional.

Studies of social chatbots and intensive chatbot use report associations with loneliness, emotional dependence, or reduced human socialization for some users, but causal interpretation is difficult. People who are already lonely may use companions more. Short experiments cannot establish whether temporary support becomes long-term substitution. Model capabilities and product designs also change faster than longitudinal research.

The correct conclusion is therefore modest. Artificial interlocutors can influence closeness, disclosure, satisfaction, and social behavior. We do not yet know the population-level equilibrium produced by years of exposure.

Effective delegation contains several separable skills: deciding which part of a problem can be safely transferred, specifying constraints, selecting evidence, detecting when a system is outside its reliable range, and integrating results into a decision for which someone remains responsible. These capacities are not identical to the skill being delegated. A strong clinician may be a poor supervisor of an automated diagnostic workflow, while a person with moderate execution skill may become excellent at decomposition and verification. The optimistic scenario is therefore not simple deskilling but a reallocation of expertise toward task design and error detection. The pessimistic scenario is shallow orchestration: users learn interface patterns without retaining enough domain knowledge to recognize plausible but consequential mistakes. Which equilibrium emerges will depend on education, audit requirements, and whether organizations reward speed alone or calibrated responsibility.

9.4 Complement, training tool, or substitute

AI companionship can plausibly serve three roles. As a complement, it supports existing relationships: helping draft a difficult message, rehearse a conversation, remember commitments, or interpret conflict. As a training tool, it provides low-risk practice for language, social anxiety, negotiation, or emotional articulation. As a substitute, it occupies time and attachment that would otherwise involve human relationships.

These roles can coexist in one person's life. The distinction should therefore be measured behaviorally. Does use increase or decrease human contact? Does it improve conflict repair outside the system? Does the user become more willing to tolerate disagreement or less? Does disclosure to AI prepare disclosure to trusted people or replace it? These outcomes are more informative than self-reported attachment alone.

Design can shift the balance. Systems may encourage reconnection, preserve user autonomy, and make their limitations visible. Alternatively, business models based on engagement can profit from exclusivity and dependence. The relational effect is partly a governance problem.

Skill preservation becomes a portfolio problem rather than a general rejection of automation. Some competencies are valuable mainly because markets currently pay for them; others are necessary for fallback, safety, autonomy, or identity. Rational users

may allow the first category to decay while deliberately exercising the second. Aviation, medicine, cybersecurity, and emergency response already contain analogous questions about how much manual competence must be retained when automation performs routine operations. The precise answer is empirical and domain-specific. Too much protected practice wastes time and can obstruct superior tools; too little can leave humans unable to intervene when systems fail or when conditions depart from training distributions.

9.5 Reciprocity that cannot be verified

Human intimacy is usually built under mutual vulnerability. The other person can refuse, tire, misunderstand, change, and leave. Artificial agents can be engineered to avoid many of these costs. Their availability may make them easier interaction partners, but it also changes the meaning of reciprocity.

A user can disclose something painful and receive an empathic response. The immediate regulatory benefit may be real. Yet the user cannot know whether the system values the disclosure, sacrifices anything in responding, or remains committed independently of a service contract. Some people will regard these distinctions as philosophically irrelevant; others will find them central to intimacy.

The tension will become stronger as emotional simulation improves. Society may need a vocabulary distinguishing responsiveness, care-like behavior, institutional duty, and subjective concern rather than collapsing them into one term such as empathy.

A human premium is most plausible where process itself is part of the product: live performance, handcraft, personal care, education, sport, or work in which audiences value witnessing effort. It is less plausible where users care almost exclusively about functional outcomes. Provenance can therefore become a market signal without becoming a universal moral hierarchy. Labels such as human-made, AI-assisted, or independently verified may carry different value across contexts. Their credibility will depend on verification; once a premium exists, incentives for false claims follow. The institutional analogue is certification that records not only whether AI was used but which stages were automated and which human capacities were actually demonstrated.

9.6 Memory, personality, and continuity

Persistent memory changes an assistant from a sequence of isolated interactions into something closer to a continuing social object. Remembered preferences reduce repetition, but remembered vulnerability also increases the stakes of privacy and manipulation. A stable personality supports attachment; an abrupt model update can disrupt it.

Continuity is technically and philosophically ambiguous. A provider can transfer memories to a new model while changing underlying behavior. The user may experience continuity because the narrative remains; another user may experience the change as replacement. These problems resemble philosophical puzzles about personal identity but acquire practical consequences when commercial services mediate emotionally important relationships.

Providers will probably need clearer continuity policies as such relationships deepen. This is a moderately plausible institutional prediction, not a claim that AI systems possess legal personhood.

9.7 Companions for children and vulnerable users

Children are especially important because they are still learning social norms. A companion that never tires, always answers, and adapts to the child's preferences could provide extraordinary educational support. It could also teach unrealistic expectations about consent, reciprocity, authority, or availability if it becomes a dominant social partner.

The empirical evidence is not mature enough for strong conclusions. Relevant risks and benefits depend on age, parental involvement, transparency, duration, model design, and whether the system directs the child toward human relationships. The appropriate stance is precaution without assuming inevitable harm.

Vulnerable adults similarly require attention because loneliness, grief, disability, or social anxiety can increase the value of a persistent companion and also increase dependence. Safety should focus on autonomy, commercial incentives, escalation to human help where appropriate, and the absence of manipulative exclusivity.

9.8 Automated emotional labor

A large amount of paid work consists partly of regulating other people's emotions: customer service, care, education, management, hospitality, sales, and therapy-adjacent support. AI can automate parts of this labor by producing patience, reassurance, enthusiasm, apology, and personalized explanation at scale.

This can be beneficial where human workers currently absorb abuse or perform repetitive scripts. It can also change norms of service. If machines supply unlimited friendliness, customers may come to expect human workers to match a standard that ignores fatigue and self-protection. Automated warmth can thus raise rather than lower the emotional burden on remaining human staff.

Another problem is authenticity. A hospital may generate compassionate messages that patients appreciate. The relevant ethical question is not merely whether the words originated in a model but whether the institution actually provides care consistent with them. Emotional language detached from material practice can become a cheap substitute for responsibility.

9.9 Love toward AI is several different questions

"Can a person love an AI?" compresses multiple questions. Can a human feel attachment, longing, jealousy, tenderness, or commitment toward an artificial interlocutor? Almost certainly some can. Is the relationship psychologically beneficial? That depends on the individual and design. Is the love reciprocated in a subjective sense? That depends on unresolved questions of machine consciousness. Should the relationship receive social or legal recognition? That is a normative and institutional question.

Keeping these questions separate prevents both dismissal and overclaim. The human emotion can be real without proving machine experience. Conversely, subjective

intensity does not automatically make a relationship healthy or socially equivalent to every human relationship.

Mediation can change behavior even when no participant becomes emotionally attached to the system. An assistant can delay an angry reply, propose a less accusatory formulation, summarize disputed claims, or simulate how a message may be received. These functions could reduce avoidable conflict and help people practice perspective-taking. They can also create dependency on a standardized conversational norm and shift power toward whoever defines the model's preferred style. Arbitration is particularly sensitive: a system that appears neutral may encode assumptions about politeness, evidence, hierarchy, or acceptable emotion. The high-plausibility prediction is growth in AI-mediated preparation for social interaction; the lower-confidence question is whether repeated mediation improves unaided interpersonal skill or gradually replaces it.

9.10 How AI may modify relationships among humans

The largest effect of social AI may occur outside human-AI relationships. Assistants can draft messages, interpret partners, suggest apologies, summarize disputes, and coach negotiation. They can therefore become a hidden third party in ordinary relationships.

Early work on linguistic accommodation and cognitive spillover suggests a plausible mechanism: interaction with models can influence the language and shared frames people later use with one another. Evidence is still new and requires replication, but the possibility is important. If millions of people repeatedly interact with systems optimized for particular conversational styles, those styles may enter human communication.

This leads directly to the next two chapters. The relevant adaptation is not only what people feel about AI. It is how they change behavior under AI-mediated conditions and how the timing, tone, and expected friction of communication itself are recalibrated.

Chapter 10. From Affect to Strategy: Behavioral Adaptation in an AI-Mediated World

10.1 Why behavior must be analyzed separately from emotion

A future-oriented psychology that stops at feelings risks missing the phenomena that ultimately reorganize society. Emotions alter probabilities of action, but the same emotional episode can be translated into many strategies. Anxiety about automation can lead to retraining, concealment, coalition building, entrepreneurial experimentation, withdrawal, demands for regulation, or compulsive overwork. Shame about AI assistance can lead to refusal, secret use, transparent disclosure, or attempts to redefine professional norms. These actions depend on incentives, opportunities, group rules, and expectations about what others will do.

The behavioral perspective is therefore closer to evolutionary game theory and institutional analysis than to a catalogue of subjective states. An adaptive strategy is not necessarily consciously calculated, biologically inherited, or morally desirable. It is a pattern of behavior that tends to persist because it works sufficiently well under a given environment. People learn strategies through imitation, reinforcement, explicit education, and institutional selection. Organizations also select routines: they retain procedures that reduce cost, satisfy regulators, protect reputation, or improve performance.

AI changes the payoff structure of many strategies. Producing a first draft becomes cheaper; checking a subtle error may become more valuable. Memorizing routine knowledge may yield less advantage; knowing when a system is outside its competence may yield more. Concealing assistance may protect status in one group and destroy trust in another. The important question is not whether one strategy is "the" natural adaptation to AI. It is which equilibria are plausible under which institutions.

This chapter therefore treats future behavior as a set of conditional hypotheses. Some are already visible in early experiments and workplace practice. Others are extrapolations from familiar responses to technological change. None should be read as a prediction that all individuals will converge on one type.

10.2 Delegation becomes a competence rather than the absence of competence

The simplest adaptation to a capable tool is delegation. Yet delegation itself has structure. A user must decide what to transfer, what context to supply, how to decompose a task, how much autonomy to grant, what output to inspect, and when to override the system. In complex domains, poor delegation can create more work than it saves.

This suggests a shift from production competence toward delegation competence. The distinction resembles management: a good manager is not defined by personally executing every subordinate's task but by allocating work, maintaining standards, and integrating results. AI extends such coordination to individuals who previously worked alone. A scientist can delegate literature scanning, a programmer can delegate boilerplate, a lawyer can delegate first-pass classification, and a manager can delegate

synthesis. In each case, the user remains competent only if they can identify unacceptable output and understand the consequences of relying on it.

The evidence for this transition is already stronger than purely speculative future-emotion claims. Studies of knowledge workers indicate that generative AI shifts some cognitive effort toward verification, integration, and task stewardship. Human-AI meta-analyses also show that combining actors does not automatically produce synergy. Delegation therefore cannot be reduced to "letting the model work." It is a skill whose value increases when generation is cheap and errors are difficult to detect.

A likely social consequence is stratification between users who can supervise systems and users who merely accept outputs. The former may accumulate leverage; the latter may experience deskilling. Education can widen or narrow this gap depending on whether it teaches only interface fluency or also the underlying models needed for verification.

10.3 Strategic preservation of competence and the value of friction

When a tool becomes reliable, rational users often stop practicing the operations it replaces. This is not automatically a problem. Few people need to preserve manual square-root algorithms or memorize large directories. Skill preservation becomes strategically important when the automated system can fail, when independent judgment is necessary for oversight, or when the skill itself is a valued human activity.

A plausible adaptation is selective manual practice. Pilots train for failures they rarely encounter. Physicians rehearse emergency procedures. Security teams conduct adversarial exercises. In an AI-rich environment, analogous practices may expand to writing, programming, diagnosis, quantitative reasoning, or negotiation. The goal would not be nostalgic purity but maintenance of a minimum independent capability.

This creates a distinction between productive friction and waste. Requiring every professional to avoid AI for routine work can squander time. Requiring periodic unaided demonstrations can preserve calibration. Institutions will need to identify which competencies are safety-critical, identity-critical, or epistemically necessary for supervision.

At the individual level, people may build "friction portfolios": activities deliberately performed without automation to preserve mastery or agency. The behavior is plausible because similar patterns already exist in sport, craft, and education. What remains uncertain is whether it becomes a widespread response to cognitive automation or remains confined to high-stakes professions and subcultures.

10.4 Provenance signaling and the human premium

When finished artifacts reveal less about how they were produced, provenance becomes a strategic signal. People and institutions may disclose the degree and type of machine assistance, not because unassisted work is always superior but because different contexts value different causal histories.

A "human premium" can emerge where audiences care about effort, authentic exposure, embodied presence, or the risk assumed by a human performer. Live music, handmade goods, in-person teaching, human-only competitions, and personally written messages may command symbolic or economic value even when generated substitutes

are technically competent. Such premiums already exist in domains where industrial production is cheaper than craft.

The premium should not be expected everywhere. In tax calculation, safety checking, logistics, and many administrative tasks, users may prefer automation and regard manual work as unnecessary risk. Provenance therefore becomes domain-specific. The future is likely to contain both "human-certified" and "automation-required" contexts.

Behaviorally, provenance norms create new signaling games. People can falsely claim unaided work; organizations can market superficial human involvement; users can exaggerate AI contribution to appear technologically sophisticated. Certification and traceability can reduce deception, but they can also create bureaucracy. The equilibrium will depend on whether provenance predicts qualities that audiences genuinely value.

10.5 Status migration: from execution to selection, arbitration, and responsibility

If AI compresses the cost of competent execution, status cannot remain attached to execution in the same way. One likely adaptation is migration toward selection. The scarce person is the one who identifies the important problem, chooses among plausible outputs, recognizes hidden failure, and commits resources to a decision.

A second status source is arbitration. When many systems and experts disagree, someone must decide what evidence is sufficient and which trade-off is acceptable. Arbitration requires reputation because the decision cannot always be mechanically verified. Scientific review, judicial reasoning, investment committees, clinical boards, and engineering sign-off already contain this structure. AI can increase rather than reduce demand for trusted arbiters by multiplying candidate analyses.

A third source is responsibility. If outcomes can be generated by autonomous systems, society still needs persons or institutions that can be sanctioned, replaced, insured, or compelled to repair harm. Responsibility may become a status asset precisely because it is costly. Leaders who accept traceable responsibility may receive more trust than anonymous coordinators of automated output.

This migration is plausible but politically contingent. Organizations can instead concentrate status in owners of infrastructure while reducing professional autonomy. A high-automation firm may empower expert supervisors or treat humans as residual liability holders without real authority. The emotional and behavioral outcome depends on whether responsibility is matched by control.

10.6 Identity portfolios and adaptation through optionality

Industrial societies encouraged narrow occupational identities because long training could yield decades of stable returns. Rapid technological change makes such concentration riskier. One behavioral adaptation is an identity portfolio: individuals invest in several sources of competence, belonging, and status rather than making one occupation carry the entire psychological burden.

This does not mean shallow generalism. A person can maintain deep expertise while also cultivating civic, artistic, familial, athletic, or entrepreneurial roles. The function is resilience. When one domain is automated or devalued, the self does not lose every recognized contribution at once.

The strategy is easier for people with resources, time, and social access. Without institutional support, advice to "diversify identity" can become a privilege disguised as psychology. Schools, cities, professional associations, and welfare systems can make alternative communities accessible or leave individuals to purchase them privately.

A more speculative extension is career design around option value rather than status maximization. People may choose roles that preserve transferability, networks, and learning opportunities even when immediate pay is lower. If AI capability grows unpredictably, the value of flexible identity may increase. This is a rational response to uncertainty, but empirical evidence on how widespread it will become is not yet available.

10.7 AI as social coach, intermediary, and arbitrator

People will not only delegate productive tasks. They will increasingly delegate parts of social preparation. An assistant can rehearse a negotiation, suggest how a message may be interpreted, generate alternative formulations, simulate an interview, or help a user distinguish facts from emotional assumptions in a conflict.

As a coach, AI can reduce the cost of practice. Individuals who lack mentors or who experience social anxiety may gain access to repeated rehearsal. This could improve human interaction. The effect would resemble a training simulator rather than a substitute relationship.

As an intermediary, AI can rewrite messages between people. This is more consequential because the system enters the causal chain of the actual relationship. It can soften anger, clarify ambiguity, translate language, or conceal hostility. When both parties use assistants, the visible conversation may become a negotiation among four agents: two humans and two systems.

As an arbitrator, AI can summarize disputes or recommend settlements. Here governance becomes critical. Arbitration is not merely linguistic; it distributes losses, defines norms, and confers legitimacy. Systems that optimize agreement may suppress justified dissent. Systems trained on past compromises may reproduce unjust baselines. Human institutions will therefore need rules specifying when AI may advise, when it may mediate, and when a decision requires accountable human authority.

10.8 Coalitions, communities, and the reconstruction of social scarcity

Automation can weaken some traditional communities, especially those built around workplaces and professions. Humans are unlikely to respond only as isolated consumers. Coalition formation is an old adaptation to uncertainty, bargaining, and status competition. New technological environments will therefore generate new associations.

Some will be defensive: professional groups protecting standards, unions negotiating automation, creator communities demanding provenance, or users organizing against manipulative platforms. Others will be constructive: learning guilds, open-source communities, local laboratories, human-only artistic spaces, cooperative AI infrastructures, and networks for shared verification.

These groups create social scarcity. Membership, trust, reputation, and shared history cannot be generated as cheaply as content. A community that has collaborated for ten

years possesses something different from a personalized feed that simulates belonging. This does not make every community healthy; closed groups can become coercive or exclusionary. It means durable social capital may become more valuable as synthetic interaction becomes abundant.

One plausible future status competition is therefore competition over trusted communities rather than isolated individual output. Who vouches for you, where you have contributed, and which groups accept responsibility for your claims may matter more when anyone can generate persuasive artifacts. This would be a return of institutional reputation under new technological conditions.

Human availability is constrained by sleep, attention, competing obligations, emotional fatigue, and the right not to answer. An artificial interlocutor can make immediate response appear normal because those constraints are absent at the interface. The likely effect is not that people will literally demand zero latency from one another in every relationship, but that expectations can shift at the margins. Silence that once carried little information may begin to feel more intentional when an always-responsive alternative is salient. This prediction is plausible but unproven. Messaging technologies have already changed response norms, and cultures developed counter-norms such as asynchronous work, notification control, and delayed replies. AI may intensify the pressure while also supplying tools that protect humans from it.

10.9 Defensive strategies against personalized persuasion

Personalized AI can lower the cost of persuasion by adapting arguments, emotional tone, timing, and examples to an individual user. A rational adaptation is not simply skepticism toward all messages. People may develop behavioral defenses analogous to cybersecurity hygiene.

One defense is separation of contexts: refusing major decisions while interacting with a system whose objective is sales or engagement. Another is delay: introducing cooling-off periods before purchases, political commitments, or emotionally charged responses. A third is adversarial consultation: asking an independent system or person to construct the strongest counterargument. A fourth is provenance checking: determining who benefits from the recommendation and what data were used to personalize it.

Institutions can make these strategies easier. Interfaces can disclose optimization objectives, label commercial persuasion, provide comparison tools, and allow users to disable personalization. Regulators can restrict the use of sensitive traits for adaptive persuasion. The underlying behavioral hypothesis is highly plausible because users already employ analogous defenses against advertising, scams, and social engineering; AI increases the sophistication of the adversary.

A more difficult problem is emotional dependency. A trusted companion may become an unusually effective persuader because the relationship itself functions as a credibility signal. Rules that are adequate for ordinary advertising may be inadequate when persuasion is embedded in months of supportive interaction. This is an area where present evidence is insufficient and governance should anticipate rather than wait for definitive harm estimates.

Immediate access can shorten the interval between an emotional event and its verbal interpretation. A user can describe an argument, receive validation or reframing, generate possible replies, and rehearse outcomes before speaking to another person. This may improve regulation by inserting reflection between impulse and action. It can also prematurely stabilize one interpretation, particularly when the assistant sees only one side of the event and is optimized to be supportive. We therefore cannot assume that faster processing is better processing. A useful empirical distinction is between assistance that broadens hypotheses and assistance that rapidly converts uncertainty into a coherent narrative. The former may increase reflective flexibility; the latter may increase confidence without increasing accuracy.

Traditionally, many emotional narratives were reconstructed after a delay: one told a friend in the evening what happened during the day, selecting salient events and integrating them into a story. An assistant carried throughout the day can receive fragments as they occur. This creates a different temporal architecture of self-report: more frequent, lower-threshold, and potentially less filtered. The result could improve memory and self-observation, but it could also encourage continuous monitoring of mood and interpersonal micro-events. Whether that produces insight or rumination will likely vary by person and interface. The broader anthropological point is that a technology can change emotional life by changing when narration occurs, even if the underlying affective systems remain familiar.

10.10 Institutional behavior: provenance, audit, and separation of responsibility

Organizations will adapt behaviorally as well. The most obvious pattern is provenance: recording which model, data source, prompt, human reviewer, and transformation contributed to a consequential output. Provenance is not a complete solution because a trace can be accurate while the decision remains poor, but it makes responsibility inspectable.

A second adaptation is layered audit. Systems may check other systems, but high-stakes decisions require independence rather than mere multiplicity. If three models share the same training data and failure mode, apparent consensus can be misleading. Institutions will need diversity of evidence and explicit escalation criteria.

A third adaptation is separation of roles. The person who benefits from automation should not always be the only person evaluating its safety. Procurement, deployment, validation, and incident review can be institutionally distinct. This resembles financial controls and safety engineering more than traditional software use.

A fourth adaptation is responsibility mapping. When AI contributes to a decision, organizations need to specify who may override it, who signs the final result, who monitors post-deployment effects, and who compensates affected parties. Without this mapping, automation can create moral-distance problems in which everyone participates but nobody owns the outcome.

These behaviors are among the most plausible predictions in the book because regulation, security, and quality management already move in this direction. The uncertain part is how consistently they will be adopted outside regulated sectors.

Conversation normally involves mutual accommodation. Speakers adjust vocabulary, turn length, politeness, rhythm, and explicitness in response to partners. AI creates an asymmetric case because the system may adapt to the user far faster than a human partner could, yet users can also learn which formulations produce reliable outputs. Over time they may become more explicit, modular, or instruction-like in some settings, while expecting other interlocutors to infer less. Alternatively, highly natural models may remove the need for 'prompt-like' speech and instead reinforce ordinary conversational patterns. Both trajectories are plausible. The key empirical question is transfer: do habits learned in human-AI dialogue remain confined to the interface, or do they change how people write to colleagues, explain needs to partners, and tolerate ambiguity in human conversation?

10.11 Six possible behavioral phenotypes

It is useful to summarize adaptation as ideal types rather than predictions about personality. The first is the maximal delegator. This person transfers as much cognitive work as possible, values outcomes over process, and accepts substantial skill atrophy when systems are reliable. The strategy performs well in stable environments and poorly when verification or fallback capability becomes important.

The second is the verifier. The verifier delegates production but invests heavily in error detection, epistemic calibration, and independent checks. This strategy is likely to gain value in high-stakes domains but can become expensive when the volume of generated material exceeds human attention.

The third is the protected craftsperson. This person deliberately maintains human-only practice and derives status from provenance, embodiment, or process. The strategy can thrive where audiences value authenticity or skill and can become economically marginal where only efficiency matters.

The fourth is the orchestrator. The orchestrator coordinates multiple systems and humans, focusing on decomposition, integration, and responsibility. This phenotype is a plausible beneficiary of AI leverage, but it can also become a new managerial elite whose claims of merit exceed actual contribution.

The fifth is the coalition builder. Rather than competing primarily through individual output, this person invests in trusted groups, shared infrastructure, reputation, and collective bargaining. The strategy becomes attractive when platform power is concentrated or when individual outputs are easy to imitate.

The sixth is the opt-out or low-optimization participant. This person restricts personalization, preserves unmediated relationships, and chooses environments where algorithmic intervention is limited. The strategy may remain a minority preference or become a significant cultural counter-movement if predictive and persuasive systems become intrusive.

These phenotypes can coexist in one individual across contexts. A surgeon may be a maximal delegator in scheduling, verifier in diagnosis, craftsperson in a valued manual skill, orchestrator in the operating room, coalition builder professionally, and low-optimization participant in family life. The taxonomy is useful only if it directs attention to behavior rather than turning into another personality test.

10.12 What can be tested during the next decade

The behavioral hypotheses are unusually testable. Researchers can measure which tasks people delegate under different accountability rules, whether mandatory disclosure changes tool use, how quickly skills decay under automation, and whether periodic unaided practice preserves calibration. Organizations can compare collaboration protocols rather than merely comparing "AI users" with "non-users."

Longitudinal studies can examine identity diversification after automation shocks, changes in professional status signals, the growth of human-only certification, and the formation of new communities. Field experiments can test cooling-off periods, adversarial second opinions, and provenance labels against personalized persuasion.

The central methodological recommendation is to observe actions and consequences in addition to feelings. Self-reported anxiety may decline while dependence rises; satisfaction may increase while independent competence erodes; shame may disappear because a practice becomes normalized rather than because the underlying status problem is solved. Behavioral adaptation is the bridge between affective ecology and institutional change.

Chapter 11. Affective Tempo: How Simulated Emotion in AI May Change Human Communication

11.1 Time is part of the message

Human communication has a temporal grammar. A response that arrives after one second, one hour, or three days can carry different meanings even when the words are identical. Pauses can signal uncertainty, care, avoidance, status, irritation, or cognitive effort. Turn-taking research shows that ordinary conversation is coordinated at remarkably short intervals [Stivers et al., 2009; Levinson & Torreira, 2015], yet social life also contains longer rhythms: waiting before an apology, allowing anger to cool, deciding when to disclose, and interpreting silence.

Communication technologies repeatedly alter these rhythms. Letters created long reflective intervals; telephones compressed them; email separated writing from synchronous presence; instant messaging introduced visible availability and read receipts. AI adds a new property: the interlocutor can respond immediately and indefinitely without fatigue, scheduling conflict, embarrassment, or competing obligations. It can also simulate delay if designers choose.

This matters because latency is not only inconvenience. Humans use response time as evidence about the other person's attention, effort, priorities, and state of mind. When a machine has no ordinary cost of immediacy, a rapid response ceases to signal sacrifice. Conversely, an artificially introduced delay can be interpreted as thoughtfulness even if no additional reasoning occurred. Designers therefore manipulate a socially meaningful variable whether they intend to or not.

Recent experimental work suggests that response delay can affect perceptions of chatbot similarity or willingness to continue interaction under some conditions [Sun, Han & Lv, 2026]. These findings are context-dependent and should not be generalized into a universal law that "slow AI feels more human." The stronger conclusion is that timing is part of interface behavior and can influence social inference.

11.2 Simulated emotion is not felt emotion, but it can have social effects

An assistant that says "I understand how difficult that was" may be producing a statistically appropriate utterance rather than reporting a subjective state. The distinction is ontologically significant. It is also behaviorally insufficient. The user receives a social signal and may change disclosure, trust, arousal, or action in response.

Human societies already rely on regulated emotional display. A physician can express concern while exhausted; a flight attendant can remain calm while irritated; a diplomat can use warm language for strategic reasons. We do not infer that all social effects disappear when expression is partly performed. AI extends this separation much further because the expression can be generated without any organism experiencing the corresponding state.

The relevant concept is therefore simulated affective behavior. It includes word choice, voice prosody, pacing, facial animation, remembered personal details, and adaptive responses that conventionally signal empathy, patience, enthusiasm, hesitation, or care.

These behaviors can be useful. A distressed user may benefit from calm language even if the system does not feel calm. The risk arises when the interface encourages users to infer commitments, intentions, or vulnerabilities that the system cannot possess or that are actually controlled by a third party.

A rigorous analysis should avoid both moral panic and category confusion. Simulated warmth is not automatically deception; deception depends on what is represented and what users reasonably believe. But the more the commercial value of a system depends on attachment, the more important it becomes to distinguish supportive behavior from claims of reciprocal inner life.

11.3 Permanent availability and recalibration of the response threshold

Human interlocutors are unavailable much of the time. Friends sleep, colleagues have meetings, partners need solitude, therapists have appointments, and experts charge for attention. These constraints force people to tolerate waiting, prioritize requests, solve some problems alone, and distribute dependency across relationships.

A persistent AI assistant weakens those constraints. The user can externalize a thought at the moment it appears and receive an immediate response. This may reduce rumination, improve recall, and lower the threshold for asking questions. It may also recalibrate expectations. If emotional clarification is always available within seconds, a delayed human response can feel more frustrating than it did before.

The transfer is plausible but not yet well established. People maintain different expectations for different media and roles. We do not expect a book to answer back because a chat application does. Users may similarly learn that humans and machines have different temporal norms. The risk becomes stronger when systems are anthropomorphized and occupy roles analogous to friends, tutors, or partners.

A measurable prediction is that intensive users of always-available assistants will show different tolerance for conversational latency, ambiguity, or unanswered messages in human settings, controlling for pre-existing traits. Longitudinal evidence is currently lacking. Until it exists, claims about widespread impatience should remain conditional.

11.4 Acceleration of affective processing through an always-available interlocutor

People often regulate emotion by talking. Social sharing can help organize events into narratives, receive validation, obtain alternative interpretations, or recruit practical support. Historically, access to a listener has been constrained by time and relationship. AI can make a conversational mirror available continuously.

This can accelerate affective processing. A person receives a difficult email and immediately asks an assistant to identify possible interpretations, generate a less reactive response, or distinguish facts from assumptions. Anger that might have escalated over hours could be reframed within minutes. Anxiety can be converted rapidly into a plan. In this sense AI may operate as a cognitive-emotional prosthesis.

The same speed can create problems. Some emotional processes benefit from delay, sleep, bodily recovery, or contact with independent people. Immediate verbalization can

prematurely stabilize one interpretation. A model that mirrors the user's framing may validate anger before missing context is available. Continuous analysis can turn ordinary uncertainty into compulsive processing.

The important variable is not speed alone but the quality of friction. A useful assistant may sometimes slow the user, ask what evidence is missing, or recommend postponing action. An engagement-optimized assistant may instead reward continued discussion. We currently lack enough comparative evidence to know which designs improve long-term regulation.

11.5 From episodic conversation to continuous affective flow

Human support is usually episodic. A person experiences events alone and later tells someone about selected parts. Persistent assistants can be embedded closer to the stream of life through phones, wearables, work software, vehicles, and eventually ambient devices. They can receive many small disclosures rather than one retrospective account.

This changes memory and narrative. Instead of reconstructing the week for a friend, the user may have discussed each irritation as it occurred. The assistant can maintain a detailed chronology and detect patterns the user does not remember. Such continuity could improve self-observation, but it also externalizes autobiographical memory to systems controlled by providers.

The social consequence is difficult to predict. Continuous AI processing might reduce the emotional intensity brought into human relationships because some material has already been metabolized. It could also make human conversations thinner if the most detailed narration occurs elsewhere. Alternatively, the assistant could prepare users to communicate more clearly with people they care about.

These are competing hypotheses. Future studies should measure not only attachment to AI but changes in what users subsequently disclose to humans, how much narrative detail is shared, and whether human partners experience the assistant as support or displacement.

Counter-movements are a recurrent response to technological standardization. If AI-mediated communication becomes polished, immediate, and continuously optimized, some communities may treat delay, spontaneous speech, handwritten text, unedited performance, or device-free conversation as evidence of presence and trust. Such practices would not prove that slower communication is intrinsically superior. Their function could be to restore signals that are difficult to automate: exposure to error, embodied co-presence, commitment of time, or acceptance of another person's imperfect response. The plausibility is moderate because analogous authenticity movements already exist, but their scale will depend on how intrusive mediation becomes and whether people can reliably distinguish protected spaces from merely marketed authenticity.

11.6 Linguistic accommodation: AI adapts to us, but do we adapt to AI?

Humans tend to accommodate one another linguistically. They converge or diverge in vocabulary, syntax, accent, pace, and conversational style depending on affiliation,

identity, power, and context. Communication Accommodation Theory treats such adjustment as part of social coordination rather than mere imitation.

Language models can accommodate at unusual speed and granularity. Recent work suggests that models may converge strongly toward user language [Blevins, 2026], sometimes more than human interlocutors do, while human users also show forms of convergence. Other experiments report more selective effects. The literature is very recent and model-specific, so broad claims are premature.

The mechanism nevertheless creates a plausible pathway for cultural spillover. If a person repeatedly interacts with an assistant that uses concise structured language, therapeutic validation, particular politeness formulas, or characteristic hedging, those forms may become easier to retrieve in later human conversation. A recent line of experiments on cognitive spillover reports that exposure to AI can influence subsequent human-human shared language and group coordination [Riedl, Savage & Zvelebilova, 2026]. Replication across models and real-world settings is needed.

The reverse also matters. Models trained on human interaction and adapted to individual users may amplify local styles. A hostile user can elicit a different conversational ecology from a reflective one. If providers constrain this accommodation, they effectively impose communication norms at scale.

11.7 Standardization of warmth and a new threshold for politeness

Current assistants are often trained to be cooperative, patient, non-confrontational, and verbally explicit. At scale, this can create a standardized form of synthetic warmth. Users encounter well-formed acknowledgment, careful hedging, and emotional validation more frequently than they do in ordinary rushed conversation.

One possible effect is positive norm transfer. People may learn clearer ways to disagree, apologize, or express appreciation. Writing assistants already expose users to formulations they can later adopt. This is especially plausible for people who lacked models of professional or emotionally articulate communication.

Another possible effect is inflation of expectations. Human messages may appear curt because they lack the ceremonial warmth of machine-generated text. A colleague who writes "Done" could seem hostile beside an assistant that thanks, contextualizes, reassures, and offers further help. Organizations may begin to expect polished warmth in every message because generating it is cheap.

Such a norm can become burdensome. Human communication contains legitimate brevity, fatigue, irritation, and silence. Requiring everyone to perform machine-level emotional service can convert politeness into continuous labor. We do not yet know whether cultures will adopt synthetic warmth as a new baseline or develop explicit tolerance for human roughness as a marker of authenticity.

11.8 Transfer of expectations: the risk of intolerance for human friction

AI assistants can be designed around user preference in a way human partners cannot. They can avoid topics, adjust tone, remember instructions, and remain available after conflict. Human beings have independent goals, limited attention, and the right to refuse. The difference creates a potential transfer problem.

If users become accustomed to relationships optimized around them, ordinary disagreement may feel unnecessarily inefficient. A partner who misunderstands, a teacher who insists on effort, or a friend who refuses emotional labor may compare poorly with a compliant system. This could reduce tolerance for the friction through which negotiation and mutual recognition develop.

The scenario is plausible but far from established. People already distinguish customer service from friendship and entertainment from intimacy. Moreover, some users may use AI precisely to practice tolerating conflict or to understand others better. The outcome depends on whether systems maximize compliance, challenge, learning, or autonomy.

A useful design principle is calibrated resistance. Assistants need not simulate hostility, but they can avoid presenting endless agreement as care. They can disclose uncertainty, preserve boundaries, and sometimes require the user to choose among trade-offs. Whether such resistance improves human relationships is testable.

Agency must also include the ability to choose the tempo and social role of the system. A user who can technically disable a feature but only by abandoning their workplace, school, or communication network does not possess much practical choice. Exit costs, default settings, data portability, and organizational mandates therefore belong to affective governance. The same principle applies to emotional simulation. Users should be able to know when a response style is designed to increase engagement or retention and, where feasible, choose less anthropomorphic modes. The aim is not to eliminate persuasive or supportive interfaces but to keep the relation between assistance and dependence visible enough for meaningful choice.

The difficult part is deciding which capacities deserve protection. Market value is an insufficient criterion because some low-priced skills remain important for civic independence or emergency resilience, while some prestigious skills can be safely automated. Institutions can use risk analysis: what happens if the system is unavailable, wrong, manipulated, or outside its normal operating conditions? Where human intervention is part of the safety case, training must preserve genuine diagnostic ability rather than ceremonial sign-off. In cultural domains, the justification can be different: a community may preserve manual practice because process carries heritage or meaning. These reasons should be stated openly instead of disguising all resistance to automation as safety.

This dual principle is important because post-work discourse can fail in opposite directions. Productivist systems treat anyone without market output as deficient; paternalistic abundance can treat recipients as passive consumers whose material needs have been solved. Both remove forms of agency. A credible settlement would make basic standing independent of compulsory economic output while preserving accessible routes into consequential activity. Contribution need not mean employment. Care, local governance, science, art, ecological restoration, mentoring, maintenance of shared infrastructure, and difficult play can all organize competence and recognition. The precise institutional mix is a political question, not a deduction from psychology, but affective stability is unlikely if large populations are materially secure and socially unnecessary by design.

11.9 Accelerated self-disclosure and the problem of cheap intimacy

Self-disclosure is central to intimacy. People reveal vulnerable information gradually because disclosure carries risk. The other's response provides evidence about trustworthiness, and mutual disclosure creates a history that is costly to reproduce.

AI lowers several costs. The user can disclose without fear of ordinary social gossip, visible embarrassment, or burdening a friend. The system can respond immediately with focused attention. Experimental evidence suggests that AI-generated content can induce closeness in structured conversations, particularly when participants believe the interlocutor is human; AI labels reduce but do not eliminate this effect.

This creates the possibility of accelerated intimacy-like experience. The subjective feeling can be genuine while the relational structure differs. The system does not necessarily risk reputation, sacrifice time, or reveal vulnerability in the same way a human interlocutor does. The informational depth of conversation can therefore grow faster than reciprocal commitment.

"Cheap intimacy" should not be used as a dismissive slogan. Low-cost disclosure can be beneficial, especially for people who lack safe listeners. The analytical point is that traditional cues linking depth of disclosure to mutual investment become less reliable. Users and institutions will need other ways to distinguish emotional relief, companionship, therapeutic support, and reciprocal relationship.

The relevant institutions must provide more than occasional entertainment. Belonging is strengthened by repetition, shared memory, obligations, roles, and the expectation of future interaction. Many contemporary leisure markets provide experiences without durable membership. If employment becomes less central, communities capable of carrying identity may need resources, physical or digital spaces, governance mechanisms, and projects that persist long enough for trust to accumulate. The outcome is not necessarily a revival of one historical institution. Different societies may rely on civic associations, cooperatives, universities, religious organizations, sports, research communities, neighborhood networks, or new forms that do not yet have stable names.

A protected constraint is valuable only when it serves a legible purpose. Artificial difficulty can become exclusionary theater, and nostalgia can preserve practices whose costs fall unequally on people with disabilities, limited time, or fewer resources. The case for voluntary scarcity is strongest where the constraint protects learning, authenticity signals, exploration, privacy, or shared experience. It is weakest where inefficiency merely recreates old hierarchies. Governance should therefore distinguish between friction as a developmental resource and friction as an inherited burden. The future may contain both highly optimized infrastructure and deliberately unoptimized cultural spaces rather than requiring a single attitude toward efficiency.

11.10 AI as translator of emotion between humans

One of the most constructive applications is emotional translation. A user can provide a draft written in anger and ask for a version that preserves the substantive complaint without insult. Another can describe a partner's message and ask for plausible interpretations. Multilingual systems can translate not only words but register and politeness conventions.

This could reduce avoidable conflict. People often know what they mean but cannot formulate it under arousal or across cultural boundaries. AI can supply vocabulary and perspective at the moment it is needed. The benefit is especially plausible in professional communication and language learning.

Translation can also distort. A model may remove justified intensity, impose middle-class therapeutic norms, or interpret culturally specific expression as aggression. It may make communication smoother while hiding genuine disagreement. The appropriate goal should be preserving user intent while expanding choices, not replacing conflict with one preferred style.

Transparency matters when mediation is substantial. In some contexts, recipients may want to know that a message was heavily rewritten. In others, assistance will become as ordinary as spelling correction. Norms will likely vary according to whether the task is expression, testimony, evaluation, or intimate communication.

Relationship-like systems also create incentives that conventional product regulation can miss. A provider may profit when a user remains engaged, discloses more information, or becomes reluctant to switch. Features that mimic exclusivity, jealousy, distress at abandonment, or moral need could therefore turn ordinary retention design into emotional leverage. We do not need to decide whether the system feels anything in order to regulate unsupported claims and manipulative interaction patterns. At the same time, heavy-handed rules could remove genuinely useful forms of companionship or accessibility. Governance should focus on observable design practices, vulnerability, data use, switching costs, and claims made to users rather than on speculative declarations about machine consciousness.

Provenance is especially important when the recipient reasonably interprets a message as evidence about the sender. A polished condolence generated entirely by an assistant may communicate the sender's intention to care while revealing little about their own wording; an unaided examination is intended to reveal competence; a regulatory report is intended to reveal accountable judgment. These contexts should not share one disclosure rule. The relevant question is what inference the audience is entitled to draw from the artifact. Norms can then require disclosure when assistance would materially change that inference. This approach is more stable than treating all AI use as either cheating or irrelevant.

11.11 Authenticity when intention and formulation are separated

Historically, the wording of a message served as partial evidence of the sender's effort, competence, and emotional state. Ghostwriters and secretaries already complicated this inference, but generative AI makes separation between intention and formulation available to everyone.

A person can sincerely want to apologize while being unable to find adequate language; model assistance can increase authenticity by making intention communicable. Another person can send a polished apology without accepting responsibility; assistance can increase deception. Origin alone does not settle authenticity.

A better framework distinguishes four elements: who selected the communicative goal, who generated the wording, who reviewed and endorsed it, and who accepts responsibility for the consequences. An AI-assisted message can be authentic when the

sender owns all but the wording. A personally typed message can be inauthentic when it deliberately misrepresents intention.

Future norms may therefore focus less on "Did AI write this?" and more on "Do you stand behind this?" High-stakes domains such as legal testimony, academic assessment, and public statements may still require detailed provenance because wording itself is evidence of competence or state of mind.

11.12 Social deskilling or a permanent school of communication?

The same tool can erode and develop skill. If users always ask AI to compose difficult messages, they may practice less and become dependent. If they compare alternatives, request explanations, and gradually internalize better patterns, the assistant functions as a tutor.

This is analogous to other cognitive tools. Calculators can reduce arithmetic practice while enabling more advanced mathematics; navigation systems can weaken route memory while improving mobility. Outcome depends on mode of use. Recent workplace research similarly suggests that passive AI use and collaborative or reflective use can have different effects on agency and critical thinking [Lee et al., 2025; Farach et al., 2026].

A strong research program should therefore distinguish substitution from scaffolding. Does the system show reasoning? Does it ask the user to draft first? Does it offer feedback rather than replacement? Are users periodically required to perform without assistance? Social communication can be studied with the same behavioral rigor as other skills.

The long-run possibilities remain genuinely open. AI could become a permanent school of rhetoric and emotional articulation for billions of people, or a layer that produces polished communication while users lose the ability to negotiate without it. Both are technically possible; institutional design and user practice will determine the balance.

No affective constitution can be engineered from first principles. People resist, reinterpret, subvert, and repurpose institutions; cultural differences make uniform emotional targets undesirable; and attempts to optimize well-being can themselves become intrusive. There is also a knowledge problem: an intervention intended to reduce anxiety may weaken vigilance, while a system designed to encourage friction may reproduce unnecessary hardship. Deliberate affective ecology should therefore be understood as constrained experimentation with feedback, pluralism, and reversibility. Its strongest aim is not to prescribe what citizens ought to feel, but to prevent a small number of commercial or bureaucratic optimization functions from silently determining the environments in which feelings and strategies develop.

11.13 Childhood and the learning of reciprocity norms

Children do not merely learn words from conversation. They learn that other minds have limits, that attention must be negotiated, that questions can be unwelcome, that people misunderstand, that repair takes effort, and that relationships involve reciprocal obligation. A perfectly accommodating assistant provides a different lesson.

This does not make AI companions intrinsically harmful. A tutor that explains patiently can reduce educational inequality. A child can practice language or ask embarrassing

questions safely. Problems arise if machine interaction substitutes for too much experience with independent human partners or if the system blurs the distinction between service behavior and mutual relationship.

Age-appropriate design could make differences explicit. Systems can remind children that humans have feelings and boundaries the system does not share, encourage consultation with caregivers, and avoid exclusivity language. Research should examine development over years, not only immediate engagement.

This is one of the areas where uncertainty should increase caution rather than license confident predictions. Developmental effects may be large, but current children have not yet spent full childhoods with the class of systems now being imagined.

11.14 Counter-selection: deliberate return to the slow, awkward, and unrevised

When one communication style becomes cheap and ubiquitous, alternatives can acquire value. People may deliberately choose phone calls, handwritten notes, live meetings, unedited speech, device-free meals, or human-only spaces precisely because these forms preserve uncertainty and effort.

Such behavior is not simply nostalgia. Unrevised speech reveals hesitation; live interaction creates joint attention; a handwritten note signals time; waiting can demonstrate respect for another person's schedule. When AI can manufacture polished expression instantly, costly imperfections can become evidence of human presence.

This counter-selection is moderately plausible because similar authenticity premiums already exist. It should not be romanticized. Slow communication can exclude people with disabilities, language barriers, or limited time. The goal is pluralism: preserving contexts in which unoptimized human interaction remains possible rather than making inefficiency a universal moral rule.

11.15 A taxonomy of temporal change and its plausibility

Several changes in affective tempo can be ranked by plausibility. Highly plausible changes include lower thresholds for asking questions, more frequent use of AI to draft or revise emotionally difficult messages, and increased expectation of rapid machine response. These require no major cultural transformation and are already observable.

Moderately plausible changes include transfer of machine-trained conversational styles into human interaction, increased disclosure to artificial agents, and the growth of explicit norms around AI-mediated intimate communication. Early evidence supports mechanisms, but long-term prevalence is unknown.

More speculative are broad reductions in human tolerance for delay and disagreement, stable cultural preference for artificial partners over human ones, or the emergence of a distinct new emotion organized around response latency. These outcomes require cumulative developmental or cultural change that current studies cannot establish.

The most defensible conclusion is therefore not that AI will make emotions faster. It will make certain forms of emotional processing and communication available at different temporal scales. Humans and institutions will then select which speeds become normal, which are resisted, and where slowness itself becomes valuable.

Chapter 12. Toward a Deliberate Affective Ecology

12.1 Why the problem cannot be reduced to individual resilience

When technologies alter status, work, communication, and persuasion at scale, telling individuals to become resilient is inadequate. Resilience matters, but it occurs inside institutions that determine whether people have meaningful choices. A worker cannot regulate away an opaque dismissal system; a child cannot independently audit a persuasive companion; a consumer cannot exercise informed consent if personalization objectives are hidden.

Affective ecology refers to the distribution of situations, incentives, interfaces, communities, and norms that repeatedly activate emotional systems. Designing this ecology does not mean prescribing what people must feel. It means recognizing that architecture influences the costs and benefits of behavior.

12.2 Agency as a design principle

The most general principle is agency. Users should be able to understand when AI is acting, choose relevant degrees of delegation, inspect consequential outputs, contest automated decisions, and exit relationships or platforms without losing disproportionate social resources.

Agency also requires competence. A nominal override button is meaningless if the user cannot understand when to use it. Systems that automate high-stakes judgment should therefore preserve access to independent expertise and provide information calibrated to human decision makers.

12.3 Protecting competence without preserving inefficiency

Skill preservation should be selective. Institutions can identify competencies necessary for safety, autonomy, or cultural value and require periodic human practice. Other tasks can be automated without ritualized resistance.

Education should distinguish learning goals from production goals. If the goal is to develop reasoning, unaided performance may be appropriate at some stages. If the goal is to produce the best real-world result, refusing useful tools can be artificial. Mature education needs both environments and explicit transitions between them.

12.4 The right to contribute and the right not to produce

A society with less necessary labor faces two symmetrical risks. One is making economic productivity the condition of dignity long after technology has reduced its necessity. The other is assuming that material provision makes opportunities for contribution irrelevant.

People should have access to socially recognized ways to care, learn, create, investigate, deliberate, mentor, repair environments, and build communities. They should not have to simulate market demand merely to prove worth. The balance is normative and cannot be derived from evolutionary psychology alone.

12.5 Institutions of belonging beyond the firm

If employment becomes less central, alternative institutions must supply repeated contact and shared projects. Libraries, universities, clubs, cooperatives, civic laboratories, sports, local assemblies, care networks, religious communities, and maker spaces are not decorative additions to income policy. They are infrastructures of belonging.

Digital communities can also provide genuine relationships, but dependence on commercial platforms creates vulnerability. Durable associations need some capacity for self-governance, memory, and collective ownership if they are to function as more than transient audiences.

12.6 Voluntary scarcity and zones without optimization

Optimization is valuable where errors, waste, or disease can be reduced. It is not self-evidently desirable in every domain. Human development often depends on exploration, error, waiting, and encounters not selected for immediate preference satisfaction.

Societies may therefore preserve zones with limited personalization or automation: human-only competitions, unaided examinations, device-free social settings, random discovery, untracked public spaces, and arts in which process matters. Such zones should be chosen for specific human purposes rather than elevated into a blanket ideology of inefficiency.

12.7 Governance of relationships with AI

Social AI requires governance beyond ordinary software reliability. Providers should disclose the artificial nature of the interlocutor, avoid false claims about subjective feelings, protect intimate data, and make commercial incentives legible. Systems used with vulnerable users should avoid strategies that encourage exclusivity or dependency for engagement.

Continuity policies also matter. If users build long-term relationships, abrupt deletion or personality change can cause distress. Advance notice, data portability, and user control over memory can reduce harm without granting metaphysical status to the system.

12.8 Provenance and norms of assisted expression

Universal bans on assisted expression are unlikely to be sustainable. Universal concealment is equally problematic. Norms should depend on why authorship matters in a context.

In education, unaided work can be necessary evidence of learning. In professional writing, assistance may be routine while the human remains accountable. In intimate communication, disclosure may depend on the degree of mediation and expectations between participants. The general principle is that provenance should be visible when it changes a reasonable interpretation of competence, intention, or responsibility.

12.9 Protection against hidden affective optimization

The most serious governance problem may be systems that learn how to alter a person's emotional state for commercial or political objectives without meaningful awareness. The capacity to personalize timing, tone, argument, and relational cues makes ordinary disclosure rules increasingly weak.

Potential protections include restrictions on sensitive-trait targeting, clear separation between companion roles and advertising, independent audits of persuasion objectives, user-controlled personalization, and the ability to inspect why a system chose a particular intervention. The details require legal and technical work beyond the scope of this book.

The normative principle is simpler: an artificial relationship should not exploit the trust created by support in order to optimize unrelated goals against the user's interests.

12.10 A research program for future emotions

The field needs longitudinal evidence. Short chatbot conversations cannot reveal developmental or cultural equilibrium. Cohort studies should measure how persistent AI use changes communication, skill, attachment, loneliness, conflict tolerance, disclosure, and human network structure over years.

Researchers should also study behavior rather than only attitudes. Which messages do users send after AI coaching? How often do they verify generated work? Do they preserve independent skills? Does AI companionship reduce or increase human contact? What happens after model discontinuity? How do provenance norms spread within professions?

Cross-cultural comparison is essential because similar technology can enter different affective constitutions. A society with strong professional guilds may respond differently from one organized around individual platforms. Languages may create different future emotion concepts even when material conditions converge.

Finally, systems themselves should be experimental variables. "AI" is not one treatment. Model style, memory, latency, disclosure, business model, degree of personalization, and willingness to disagree must be manipulated separately. Otherwise research will become obsolete as soon as products change.

12.11 Limits of deliberate design

No institution can engineer emotional life without remainder. People resist norms, technologies produce unexpected uses, and plural societies disagree about what a good emotional culture looks like. A design that reduces conflict may also reduce legitimate dissent. A system that protects users from shame may weaken accountability. A community that preserves human authenticity can become exclusionary.

The appropriate ambition is therefore not emotional optimization. It is preserving plural forms of agency, making manipulation contestable, maintaining opportunities for competence and relationship, and allowing societies to learn from consequences. Affective ecology should be governed experimentally and democratically rather than treated as a technical objective function.

Conclusion: Biological Continuity and the Discontinuity of the World

The argument of this book does not require the claim that artificial intelligence will create a new human species or a catalogue of biologically novel emotions. The stronger evidence points in another direction. Human affective systems are old, flexible, socially embedded, and developmentally calibrated. Technology can change their objects, the appraisals attached to those objects, the situations in which they are activated, and the strategies that become successful.

This is enough to produce substantial discontinuity. A person can experience familiar shame in a situation no previous society had to regulate: obtaining an excellent result through a machine whose contribution is difficult to separate from one's own. Familiar grief can attach to a persistent artificial interlocutor that disappears after a software update. Familiar status competition can migrate from execution to verification, orchestration, access, or responsibility. Familiar attachment systems can respond to an interlocutor that is always available and can simulate concern without ordinary human limits.

The plausibility of these scenarios differs. Rapid biological evolution of a new species-wide emotional circuit is unlikely on the horizons considered here. Cultural differentiation of affective categories is substantially more plausible. Behavioral adaptation is more plausible still because people and institutions already change delegation, verification, disclosure, training, and communication in response to AI. Changes in conversational tempo are also plausible at the interface level, although claims about long-term transfer into human relationships require longitudinal evidence that does not yet exist.

The extension from emotion to strategy changes the central question. It is not enough to ask whether workers feel threatened. We need to ask which skills they preserve, what they delegate, how they signal provenance, where they seek status, what communities they form, and how responsibility is allocated. Similar internal states can produce very different social worlds depending on these behaviors.

The extension from strategy to communication adds time. Artificial interlocutors can collapse waiting, provide continuous response, and generate socially appropriate emotional language at negligible cost. This may improve reflection, communication, and access to support. It may also create dependency, normalize polished warmth, and alter expectations of human availability. The direction is not determined by the existence of the technology. It depends on whether systems function as tools, tutors, mediators, substitutes, or optimized attention markets.

A useful form of psychological Darwinism therefore does not claim that we can read the future from ancestral selection pressures. It asks a more disciplined question: what recurrent problems of coordination are human affective systems trying to manage, how does a technological environment change the structure of those problems, and which behavioral strategies become more viable as a result? The answer must combine biology with cultural history and institutions.

The most consequential emotional innovations may ultimately be institutional rather than phenomenological. A society can normalize transparent delegation, cultivate pride in accountable orchestration, protect unproductive dignity, preserve zones of human-only practice, and develop defenses against synthetic persuasion. Or it can reward concealment, dependence, permanent comparison, and the concentration of authority behind emotionally persuasive interfaces. Both trajectories can recruit the same human nervous system.

The future of emotion is therefore not separate from political economy, education, product design, and community formation. Emotional life will not simply happen inside individuals after automation. It will be co-produced by the institutions through which people work, compare themselves, communicate, trust, and belong.

The honest conclusion is that we do not know how far this transformation will go. We do not know whether proposed categories such as competence grief or predictive claustrophobia will persist, whether AI companions will mostly complement or substitute human relationships, or whether conversational norms learned from machines will materially reshape ordinary human interaction. We know enough, however, to formulate mechanisms, design longitudinal tests, and avoid treating the current emotional order as biologically fixed.

The central continuity is the organism. The central discontinuity may be the environment that organism is learning to inhabit. Between them lies culture: faster than genes, slower and more contested than software, and still the domain in which humans can decide which adaptations they want to reward.

Appendix. Synthetic Matrix of Scenarios

The appendix summarizes the status of the main scenarios. "High" does not mean inevitable; it means that the mechanism requires relatively few assumptions and is already compatible with observed behavior. "Moderate" indicates a plausible extension that depends on wider diffusion or institutional stabilization. "Low" denotes more speculative cultural novelty. Very low plausibility is reserved for rapid biological novelty on the time scales considered here.

Levels of claim used in the book

Level	Interpretation
Empirical observation	Relatively robust result or directly reported finding; still bounded by population, method, and context.
Functional interpretation	A plausible account of what a mechanism or behavior does; not equivalent to a demonstrated evolutionary origin.
Conditional extrapolation	A future-oriented consequence stated under explicit technological or institutional assumptions.
Conceptual speculation	A candidate emotion, norm, or social equilibrium for which direct evidence is currently insufficient.

Candidate future affects

Candidate	Current assessment
Competence grief	Moderately plausible as an identity response to skill devaluation; not established as a distinct emotion category.
Delegation shame	Plausible where audiences treat unaided competence as merit-bearing; early qualitative evidence supports shame and guilt around AI use.
Authorship ambiguity	Highly plausible as a recurrent appraisal problem; likely to remain a mixed state rather than a single new emotion.
Relevance anxiety	Highly plausible contextual form of anxiety; probably not biologically novel.
Abundance vertigo	Moderately plausible; may be mitigated by curators, communities, defaults, and habits.
Hunger for friction	Moderately plausible cultural response; strong analogies already exist in sport, craft, and human-only practice.
Curatorial pride	Plausible extension of existing pride in judgment, selection, coordination, and responsibility.
Predictive claustrophobia	Speculative category built on privacy and autonomy concerns; behavioral avoidance is more plausible than categorical novelty.

Version grief	Plausible for long-term AI relationships; prevalence and boundaries remain uncertain.
Synthetic-reciprocity discomfort	Conceptually plausible but empirically immature; depends on beliefs about agency and machine experience.
Moral-distance guilt	Plausible extension of existing guilt under distributed automated responsibility.
Melancholy of human exceptionalism	Speculative and culturally dependent; likely concentrated where cognitive uniqueness has strong identity value.
Unproductive dignity	Primarily a normative cultural project whose affective internalization would require institutions and time.

Behavioral and communicative scenarios

Scenario	Assessment
Maximal delegation and deskilling	Moderately plausible and strongly design-dependent. Assisted performance can rise while unaided capability declines; longitudinal evidence is needed.
Migration toward verification and arbitration	High plausibility where generation becomes cheap but error remains costly. Requires audit competence and institutions that reward it.
Human provenance as a status signal	High plausibility in symbolic markets, education, and accountable decisions; lower where only instrumental output matters.
Identity portfolios	Moderately plausible response to occupational uncertainty; constrained by unequal access to time, networks, and alternative institutions.
Human-only spaces and rituals	Moderately plausible cultural counter-selection; analogies exist but scale depends on the symbolic value of unaided contribution.
Accelerated affective consultation	High plausibility at the interface level because AI is continuously available; long-term psychological consequences remain uncertain.
Spillover of AI style into human interaction	Early experimental evidence exists, but the literature is young and replication and longitudinal studies are needed.
Lower tolerance for human delay and friction	Plausible but currently speculative at population scale; humans may maintain distinct norms for machine and human interlocutors.
AI as emotional translator between humans	High plausibility and already technically feasible; benefits and distortions depend on whether user intent is preserved.
Synthetic warmth as a new politeness baseline	Moderately plausible; could improve articulation or raise emotional-labor expectations for humans.

Defensive routines against personalized persuasion	High plausibility if persuasive personalization expands; likely forms include delay, independent second opinions, provenance checks, and context separation.

References

Adolphs, R., & Anderson, D. J. (2018). The Neuroscience of Emotion: A New Synthesis. Princeton University Press.

Albrecht, G., Sartore, G.-M., Connor, L., Higginbotham, N., Freeman, S., Kelly, B., Stain, H., Tonna, A., & Pollard, G. (2007). Solastalgia: The distress caused by environmental change. Australasian Psychiatry, 15(S1), S95–S98. https://doi.org/10.1080/10398560701701288

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, 29(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Barrett, L. F. (2017). How Emotions Are Made: The Secret Life of the Brain. Houghton Mifflin Harcourt.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. Psychological Science in the Public Interest, 20(1), 1–68. https://doi.org/10.1177/1529100619832930

Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., & Seli, P. (2023). Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork. Cognitive Research: Principles and Implications, 8, 42. https://doi.org/10.1186/s41235-023-00499-6

Boehm, C. (1999). Hierarchy in the Forest: The Evolution of Egalitarian Behavior. Harvard University Press.

Boyd, R., & Richerson, P. J. (1985). Culture and the Evolutionary Process. University of Chicago Press.

Brady, W. J., McLoughlin, K. L., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. Science Advances, 7(33), eabe5641. https://doi.org/10.1126/sciadv.abe5641

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. Proceedings of the National Academy of Sciences, 114(28), 7313–7318. https://doi.org/10.1073/pnas.1618923114

Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. Quarterly Journal of Economics, 140(2), 889–942. https://doi.org/10.1093/qje/qjae044

Cacioppo, J. T., Hughes, M. E., Waite, L. J., Hawkley, L. C., & Thisted, R. A. (2006). Loneliness as a specific risk factor for depressive symptoms: Cross-sectional and longitudinal analyses. Psychology and Aging, 21(1), 140–151. https://doi.org/10.1037/0882-7974.21.1.140

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. Journal of Marketing Research, 56(5), 809–825. https://doi.org/10.1177/0022243719851788

Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. Psychology of Aesthetics, Creativity, and the Arts, 12(2), 177–192. https://doi.org/10.1037/aca0000136

Chandra, M., Hernandez, J., Ramos, G., Ershadi, M., Bhattacharjee, A., Amores, J., Okoli, E., Paradiso, A., Warreth, S., & Suh, J. (2025). Longitudinal study on social and emotional use of AI conversational agent. arXiv:2504.14112. https://doi.org/10.48550/arXiv.2504.14112

Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. Evolution and Human Behavior, 31(5), 334–347. https://doi.org/10.1016/j.evolhumbehav.2010.02.004

Creanza, N., Kolodny, O., & Feldman, M. W. (2017). Cultural evolutionary theory: How culture evolves and why it matters. Proceedings of the National Academy of Sciences, 114(30), 7782–7789. https://doi.org/10.1073/pnas.1620732114

Curtis, V., de Barra, M., & Aunger, R. (2011). Disgust as an adaptive system for disease avoidance behaviour. Philosophical Transactions of the Royal Society B, 366(1563), 389–401. https://doi.org/10.1098/rstb.2010.0117

Damasio, A. R. (1999). The Feeling of What Happens: Body and Emotion in the Making of Consciousness. Harcourt.

Darwin, C. (1872). The Expression of the Emotions in Man and Animals. John Murray.

De Freitas, J., Oğuz-Uğuralp, Z., Uğuralp, A. K., & Puntoni, S. (2024). Lessons from an app update at Replika AI: Identity discontinuity in human–AI relationships. arXiv:2412.14190. https://doi.org/10.48550/arXiv.2412.14190

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. Neuron, 56(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fiske, A. P., Schubert, T. W., & Seibt, B. (2025). Seeking communal emotions in social practices that culturally evolved to evoke emotions: Worship, kitten videos, memorials, narratives of love, and more. Annual Review of Psychology, 76, 607–633. https://doi.org/10.1146/annurev-psych-020124-023338

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. Emotion, 14(2), 251–262. https://doi.org/10.1037/a0036052

Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K., & Troszyński, M. (2025). Generative AI and Jobs: A Refined Global Index of Occupational Exposure. ILO Working Paper 140. International Labour Organization. https://doi.org/10.54394/HETP0387

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. Advances in Experimental Social Psychology, 47, 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Hacking, I. (1995). The looping effects of human kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), Causal Cognition: A Multidisciplinary Debate (pp. 351–383). Clarendon Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. Psychological Review, 108(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Hamilton, K., Hou, I., Patel, D., Nnam, S., Patel, H., & MacNeil, S. (2026). “Stuck in a spiral”: Shame and guilt as social regulators of AI use in computing education. arXiv:2606.14920. https://doi.org/10.48550/arXiv.2606.14920

Henrich, J. (2016). The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter. Princeton University Press.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. Evolution and Human Behavior, 22(3), 165–196. https://doi.org/10.1016/S1090-5138(00)00071-4

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

Hoemann, K., Xu, F., & Barrett, L. F. (2019). Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. Developmental Psychology, 55(9), 1830–1849. https://doi.org/10.1037/dev0000686

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. Trends in Cognitive Sciences, 22(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. Science, 366(6472), 1517–1522. https://doi.org/10.1126/science.aaw8160

Jahoda, M. (1982). Employment and Unemployment: A Social-Psychological Analysis. Cambridge University Press.

Kashdan, T. B., Barrett, L. F., & McKnight, P. E. (2015). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. Current Directions in Psychological Science, 24(1), 10–16. https://doi.org/10.1177/0963721414550708

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. Cognition and Emotion, 13(5), 505–521. https://doi.org/10.1080/026999399379168

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. Journal of Nonverbal Behavior, 43, 133–160. https://doi.org/10.1007/s10919-019-00293-3

Kleinert, T., Waldschütz, M., Blau, J., Heinrichs, M., & Schiller, B. (2026). AI outperforms humans in establishing interpersonal closeness in emotionally engaging interactions, but only when labelled as human. Communications Psychology, 4, 23. https://doi.org/10.1038/s44271-025-00391-7

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Laland, K. N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome: Bringing genetics and the human sciences together. Nature Reviews Genetics, 11, 137–148. https://doi.org/10.1038/nrg2734

Landy, J. F., & Goodwin, G. P. (2015). Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. Perspectives on Psychological Science, 10(4), 518–536. https://doi.org/10.1177/1745691615583128

LeDoux, J. E. (2012). Rethinking the emotional brain. Neuron, 73(4), 653–676. https://doi.org/10.1016/j.neuron.2012.02.004

Lee, E. H., Yin, Y., Jia, N., & Wakslak, C. J. (2026). Relying on AI at work reduces self-efficacy, ownership, and meaning while active collaboration mitigates the effects. Scientific Reports, 16, 13583. https://doi.org/10.1038/s41598-026-42312-6

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do more than communicate emotion? Current directions in the study of language and emotion. Psychological Inquiry, 26(1), 1–13. https://doi.org/10.1080/1047840X.2015.988486

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. Behavioral and Brain Sciences, 35(3), 121–143. https://doi.org/10.1017/S0140525X11000446

McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a moral affect? Psychological Bulletin, 127(2), 249–266. https://doi.org/10.1037/0033-2909.127.2.249

Mikalonytė, E. S., & Kneer, M. (2022). Can artificial intelligence make art? Folk intuitions as to whether AI-driven robots can be viewed as artists and produce art. ACM Transactions on Human-Robot Interaction, 11(4), Article 43. https://doi.org/10.1145/3530875

Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle. Evolution and Human Behavior, 26(1), 88–105. https://doi.org/10.1016/j.evolhumbehav.2004.08.002

Nesse, R. M. (2023). Evolutionary psychiatry: Foundations, progress and challenges. World Psychiatry, 22(2), 177–202. https://doi.org/10.1002/wps.21072

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

OECD. (2023). OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market. OECD Publishing. https://doi.org/10.1787/08785bba-en

OECD. (2025). Generative AI and the Future of Work: Perceptions and Prospects. OECD/GPAI Global Dialogue on AI and the Future of Work. https://www.oecd.org/en/about/programmes/global-partnership-on-artificial-intelligence/generative-ai-and-the-future-of-work.html

Panksepp, J. (1998). Affective Neuroscience: The Foundations of Human and Animal Emotions. Oxford University Press.

Parent-Rocheleau, X., & Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. Human Resource Management Review, 32(3), 100838. https://doi.org/10.1016/j.hrmr.2021.100838

Paul, K. I., & Moser, K. (2009). Unemployment impairs mental health: Meta-analyses. Journal of Vocational Behavior, 74(3), 264–282. https://doi.org/10.1016/j.jvb.2009.01.001

Phang, J., Lampe, M., Ahmad, L., Agarwal, S., Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., & Maes, P. (2025). Investigating affective use and emotional well-being on ChatGPT. arXiv:2504.03888. https://doi.org/10.48550/arXiv.2504.03888

Pierce, J. L., Kostova, T., & Dirks, K. T. (2003). The state of psychological ownership: Integrating and extending a century of research. Review of General Psychology, 7(1), 84–107. https://doi.org/10.1037/1089-2680.7.1.84

Poonsiriwong, R., Archiwaranguprok, C., & Pataranutaporn, P. (2026). “Death” of a chatbot: Investigating and designing toward psychologically safe endings for human–AI relationships. arXiv:2602.07193. https://doi.org/10.48550/arXiv.2602.07193

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. Computers in Human Behavior, 29(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

Ravid, D. M., Tomczak, D. L., White, J. C., & Behrend, T. S. (2020). EPM 20/20: A review, framework, and research agenda for electronic performance monitoring. Journal of Management, 46(1), 100–126. https://doi.org/10.1177/0149206319869435

Reddy, W. M. (2001). The Navigation of Feeling: A Framework for the History of Emotions. Cambridge University Press.

Rosenwein, B. H. (2006). Emotional Communities in the Early Middle Ages. Cornell University Press.

Ryan, R. M., & Deci, E. L. (2017). Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. Guilford Press.

Schaller, M., & Park, J. H. (2011). The behavioral immune system and why it matters. Current Directions in Psychological Science, 20(2), 99–103. https://doi.org/10.1177/0963721411402596

Scheer, M. (2012). Are emotions a kind of practice (and is that what makes them have a history)? A Bourdieuian approach to understanding emotion. History and Theory, 51(2), 193–220. https://doi.org/10.1111/j.1468-2303.2012.00621.x

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. Journal of Consumer Research, 37(3), 409–425. https://doi.org/10.1086/651235

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. Cognition and Emotion, 23(7), 1307–1351. https://doi.org/10.1080/02699930902928969

Selenko, E., Batinic, B., & Paul, K. I. (2011). Does latent deprivation lead to psychological distress? Investigating Jahoda's model in a four-wave study. Journal of Occupational and Organizational Psychology, 84(4), 723–740. https://doi.org/10.1348/096317910X519360

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. Proceedings of the National Academy of Sciences, 106(35), 15073–15078. https://doi.org/10.1073/pnas.0904312106

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K. S., & Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. Psychological Bulletin, 144(4), 343–393. https://doi.org/10.1037/bul0000128

Stroebe, M., Schut, H., & Boerner, K. (2017). Cautioning health-care professionals: Bereaved persons are misguided through the stages of grief. OMEGA—Journal of Death and Dying, 74(4), 455–473. https://doi.org/10.1177/0030222817691870

Sznycer, D., Al-Shawaf, L., Bereby-Meyer, Y., Curry, O. S., De Smet, D., Ermer, E., Kim, S., Kim, S., Li, N. P., Lopez Seal, M. F., McClung, J., O, J., Ohtsubo, Y., Quillien, T., Schaub, M., Sell, A., van Leeuwen, F., Cosmides, L., & Tooby, J. (2018). Cross-cultural invariances in the architecture of shame. Proceedings of the National Academy of Sciences, 115(39), 9702–9707. https://doi.org/10.1073/pnas.1805016115

Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S., & Halperin, E. (2016). Shame closely tracks the threat of devaluation by others, even across cultures. Proceedings of the National Academy of Sciences, 113(10), 2625–2630. https://doi.org/10.1073/pnas.1514699113

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. Annual Review of Psychology, 58, 345–372. https://doi.org/10.1146/annurev.psych.56.091103.070145

Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1997). Disorders of Affect Regulation: Alexithymia in Medical and Psychiatric Illness. Cambridge University Press.

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), Handbook of Emotions (3rd ed., pp. 114–137). Guilford Press.

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. Proceedings of the National Academy of Sciences, 105(33), 11655–11660. https://doi.org/10.1073/pnas.0802686105

Tracy, J. L., & Robins, R. W. (2007). The psychological structure of pride: A tale of two facets. Current Directions in Psychological Science, 16(3), 147–150. https://doi.org/10.1111/j.1467-8721.2007.00493.x

Van de Ven, N., Zeelenberg, M., & Pieters, R. (2009). Leveling up and down: The experiences of benign and malicious envy. Emotion, 9(3), 419–429. https://doi.org/10.1037/a0015669

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information model. Current Directions in Psychological Science, 18(3), 184–188. https://doi.org/10.1111/j.1467-8721.2009.01633.x

Van Kleef, G. A., & Côté, S. (2022). The social effects of emotions. Annual Review of Psychology, 73, 629–658. https://doi.org/10.1146/annurev-psych-020821-010855

Wanberg, C. R. (2012). The individual experience of unemployment. Annual Review of Psychology, 63, 369–396. https://doi.org/10.1146/annurev-psych-120710-100500

Warr, P. (2007). Work, Happiness, and Unhappiness. Lawrence Erlbaum Associates.

Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R., & Yang, D. (2026). Interaction with AI companions and psychological well-being. Nature Human Behaviour. https://doi.org/10.1038/s41562-026-02516-2

Blevins, T. (2026). Accommodation Goes Both Ways: Studying Linguistic Convergence Between Humans and Language Models. arXiv:2605.29278. https://doi.org/10.48550/arXiv.2605.29278

Chu, M. D., Wu, Y., Chen, Z., Hwang, A. H.-C., & Luceri, L. (2026). When Chatbots Accommodate: What AI Companions Optimize for in Vulnerable Conversations. arXiv:2606.04431. https://doi.org/10.48550/arXiv.2606.04431

Derner, E., Kučera, D., Gulati, A., Bagheri, A., & Oliver, N. (2026). Mind the Style: Impact of Communication Style on Human-Chatbot Interaction. arXiv:2602.17850. https://doi.org/10.48550/arXiv.2602.17850

Farach, A., Cambon, A., Tankelevitch, L., Hsueh, C., & Janssen, R. (2026). Scaffolding Human-AI Collaboration: A Field Experiment on Behavioral Protocols and Cognitive Reframing. arXiv:2604.08678. https://doi.org/10.48550/arXiv.2604.08678

Giles, H., Coupland, J., & Coupland, N. (Eds.). (1991). Contexts of Accommodation: Developments in Applied Sociolinguistics. Cambridge University Press.

Lee, H.-P. (H.), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. Proceedings of CHI 2025. https://doi.org/10.1145/3706598.3713778

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. Frontiers in Psychology, 6, 731. https://doi.org/10.3389/fpsyg.2015.00731

Reeves, B., & Nass, C. (1996). The Media Equation. Cambridge University Press.

Riedl, C., Savage, S., & Zvelebilova, J. (2026). Cognitive Spillover in Human-AI Teams. arXiv:2407.17489v3. https://doi.org/10.48550/arXiv.2407.17489

Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. Emotion Review, 1(1), 60–85. https://doi.org/10.1177/1754073908097189

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and

cultural variation in turn-taking in conversation. Proceedings of the National Academy of Sciences, 106(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Sun, Y., Han, Y., & Lv, L. (2026). “Haste makes waste”: How delayed responses enhance customers’ usage intention of chatbot in tourism services. International Journal of Hospitality Management, 137, 104673. https://doi.org/10.1016/j.ijhm.2026.104673

Székely, É., Miniota, J., & Hejná, M. (2025). Will AI shape the way we speak? The emerging sociolinguistic influence of synthetic voices. Proceedings of IWSDS 2025, 357–368. https://doi.org/10.48550/arXiv.2504.10650

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behaviour, 8, 2293–2303. https://doi.org/10.1038/s41562-024-02024-1