

Technology and the Virtues

Ethics in the Digital Age

Brendan Shea, PhD

Rochester Community and Technical College

November 18, 2025

What Makes a Good Life? Introduction to Virtue Ethics

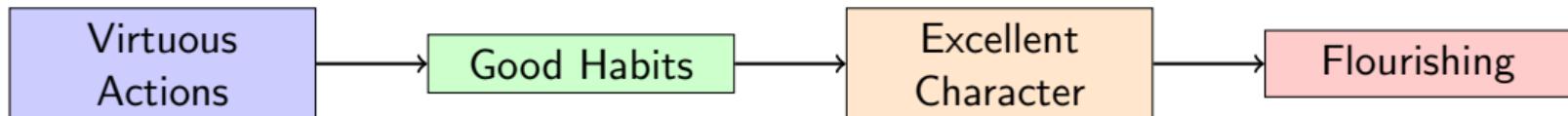
- Most ethical theories ask "What rules should I follow?" or "What consequences will my actions have?"
- **Virtue ethics** takes a different approach: it asks "What kind of person should I become?"
- Rather than focusing only on individual actions, virtue ethics examines our character, habits, and overall way of life.
- The central question becomes: "Am I developing into a person who lives well and flourishes?"

Key Definition

Virtue ethics: An ethical approach that focuses on developing good character traits (virtues) rather than following rules or calculating consequences.

Aristotle's Key Ideas: Character, Habits, and Human Flourishing

- Aristotle (384-322 BCE) argued that humans have a distinctive function or purpose: to live according to reason.
- **Eudaimonia** (often translated as "flourishing" or "the good life") is achieved by developing excellent character traits.
- Virtues are not inborn—they must be cultivated through practice, much like learning to play an instrument.
- We become virtuous by repeatedly performing virtuous actions until they become our habits and shape our character.



Modern Virtue Ethics: Vallor and Haidt's Approach

- Shannon Vallor (philosopher, University of Edinburgh) and Jonathan Haidt (social psychologist, NYU) both apply virtue ethics to modern technology.
- They argue that to evaluate technologies like social media and AI, we must ask: "Do these help or hinder human flourishing?"
- Both thinkers focus on how technologies shape our character development, especially during crucial periods like childhood and adolescence.
- Their approach is **developmental**: they examine how technology affects our ability to cultivate virtues over time.

Author	Primary Focus
Shannon Vallor	Social media ethics and AI's threat to human agency
Jonathan Haidt	Social media's impact on childhood development

Why Virtue Ethics for Technology? Beyond Rules and Consequences

- Traditional approaches ask: "Did this violate someone's rights?" (rule-based) or "Did this cause harm?" (consequence-based).
- These questions are important, but virtue ethics adds something crucial: "What kind of person is this technology helping me become?"
- Technology isn't neutral—it shapes our habits, attention, relationships, and ultimately our character.
- A virtue ethics lens reveals how seemingly small technological choices compound over time to fundamentally change who we are.

The Central Question

Do our technologies help us develop the virtues necessary for human flourishing—or do they undermine our character development?

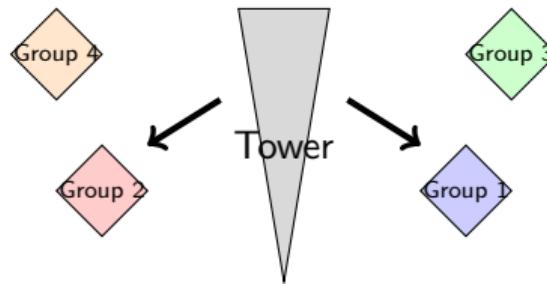
The Central Question: Do Our Technologies Help Us Become Good People?

- Vallor and Haidt share a common concern: technology is not just a tool we use, but a force that shapes who we become.
- Every time we interact with social media or AI, we are practicing certain habits and reinforcing certain character traits.
- The question is not whether these technologies are "good" or "bad" in isolation, but whether they cultivate virtue or vice in us.
- Technologies that seem harmless or even beneficial in the short term may undermine our character development over years.

"The challenge we face today is not a moral dilemma; it is rather a moral imperative, long overdue in recognition, to collectively cultivate the technomoral virtues needed to confront... emerging technosocial challenges wisely and well." — Shannon Vallor

The Tower of Babel: Haidt's Metaphor for What Went Wrong

- Haidt uses the biblical story of the Tower of Babel to describe what happened to American society around 2010-2015.
- In the story, God confuses the language of people building a tower to heaven, making them unable to communicate or cooperate.
- Similarly, social media platforms fragmented our shared reality and made genuine dialogue increasingly difficult.
- The result is a society where we inhabit "broken pieces of glass"—isolated echo chambers with no common understanding.



Vallor's Mirror: AI Reflects Us But Doesn't Think

- Vallor argues that AI systems like ChatGPT are not minds—they are **mirrors** that reflect human intelligence back at us.
- Just as a bathroom mirror shows your face but doesn't have warmth or depth, AI produces **reflections of intelligence** without understanding or experience.
- The danger is not that AI will become conscious and rebel, but that we will mistake these reflections for real thinking.
- When we treat AI as a mind, we are tempted to diminish what makes human minds unique and valuable.

Key Distinction

Mind: Characterized by experience, understanding, consciousness, and meaning-making.
Mirror: Reflects the patterns of minds without possessing these qualities itself.

Agreement: Both Social Media AND AI Threaten Human Flourishing

- Both writers have written extensively on social media (and its harms) before turning to AI in recent years.
- Social media undermines our ability to form deep friendships, think critically, and participate in democracy.
- AI threatens to make us passive, reducing human agency and our confidence in human judgment.
- Both technologies share a common problem: they interfere with the experiences and practices necessary for virtue development.

Social Media Problems:

- Fragmented attention
- Shallow relationships
- Echo chambers
- Performative identity

AI Problems:

- Outsourced thinking
- Loss of meaning-making
- Reduced human agency
- Reductive view of mind

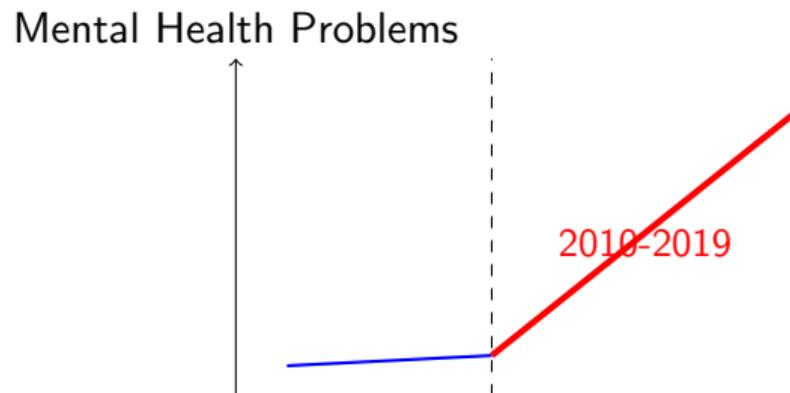
The Timing: Why 2010-2015 Was the Critical Period

- Before 2010, the internet and early social media (Myspace, basic Facebook) did not show negative effects on teen mental health.
- Between 2010 and 2015, several technological changes converged: smartphones became ubiquitous, social media added "Like" and "Share" buttons, and algorithms began prioritizing engagement.
- This created what Haidt calls **enhanced virality**—content that triggered strong emotions (especially outrage) spread faster than ever before.
- Both authors identify this brief period as when childhood and social life were fundamentally "rewired" across the developed world.

Year	Technology	Impact
2007	iPhone released	Constant internet access
2009	Facebook "Like" button	Quantified social approval
2010	Instagram launched	Visual social comparison
2012	Smartphones reach 50% of teens	Always-on connectivity

Haidt's Evidence: The Mental Health Crisis by the Numbers

- Between 2010 and 2019, rates of depression and anxiety among U.S. adolescents rose by more than 50 percent in many studies.
- The suicide rate for adolescents ages 10-19 rose 48 percent; for girls ages 10-14, it rose 131 percent.
- These patterns appeared simultaneously in Canada, the U.K., Australia, New Zealand, and Nordic countries—suggesting a common cause.
- Academic performance also declined: after decades of improvement, reading and math scores began dropping globally after 2012.



Vallor's Evidence: From Social Media to Silicon Valley's AI Dreams

- Vallor's early work (2010s) documented how social media platforms were designed to maximize **engagement**—keeping users scrolling, clicking, and reacting.
- She observed that social media was already promoting a reductive view of human relationships: reducing friendship to "connections" and social life to quantifiable metrics.
- More recently, she argues that AI development in Silicon Valley has doubled down on this reductive vision, treating human intelligence as mere "pattern matching."
- Both social media and AI embody what she calls the "ideology of efficiency"—measuring everything by optimization and ignoring deeper human values.

The Progression of Concern

First, social media reduced human relationships to data. Now, AI threatens to reduce human thinking itself to algorithmic processing.

What Virtues Require: Experience, Practice, and Risk

- Virtues are not abstract principles—they are **character traits** developed through repeated practice in challenging real-world situations.
- To develop courage, we must face genuine fears and risks; to develop practical wisdom, we must make difficult decisions with real consequences.
- Friendship requires embodied interaction where we learn to read body language, navigate conflicts, and support each other in person.
- **Civic virtues** like tolerance, compromise, and good judgment require exposure to diverse viewpoints and practice in democratic deliberation.

Three Requirements for Virtue Development

- ① **Experience:** Direct, embodied engagement with the world and others
- ② **Practice:** Repeated actions that form habits and shape character
- ③ **Risk:** Exposure to genuine consequences that make virtues meaningful

How Technology Blocks Virtue Development

- Social media and AI systematically remove the three conditions necessary for virtue development: experience, practice, and risk.
- Online interactions lack the **embodied presence** that teaches us to read emotions, navigate conflicts, and build deep trust.
- Algorithms optimize for engagement and efficiency, removing the friction and difficulty that force us to develop practical wisdom.
- When AI handles our thinking and social media mediates our relationships, we lose opportunities to practice the virtues we need to flourish.

What Virtues Need:

- Face-to-face interaction
- Genuine consequences
- Difficult decisions
- Unscripted moments
- Real accountability

What Technology Provides:

- Mediated screens
- Easy escape/blocking
- Algorithmic guidance
- Curated feeds
- Anonymous actions

Case Study: Three Generations of Friendship—What Changed?

- **Grandmother (born 1950s)**: Maintained friendships through weekly letters, occasional long-distance phone calls, and regular in-person visits.
- **Parent (born 1980s)**: Used email and early internet to stay connected, but most friendship activities still happened face-to-face.
- **Teen (born 2008)**: Primary friendship maintenance happens through Snapchat streaks, Instagram comments, and text messages; sees friends in person mainly at school.
- The teen has 200+ "friends" online but reports feeling lonely and reports only 2-3 people she could call in a crisis.

Questions to Consider

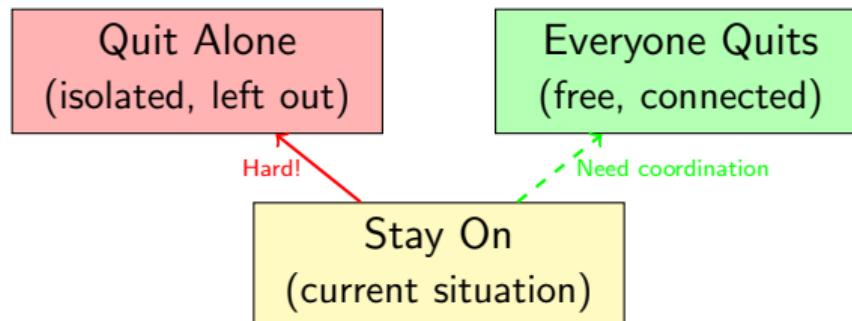
Which generation's friendships best cultivated virtues like loyalty, empathy, courage, and practical wisdom? What was gained and lost in each transition?

Discussion Questions: What Do Friendships Need to Flourish?

- ① Haidt identifies four features of real-world interaction that are diminished online: embodied, synchronous, one-to-one, and high-cost communities. Which of these do you think is most important for developing genuine friendship?
- ② Vallor argues that online friendships can be "reflections" of real friendship without the depth of actual friendship. Have you experienced a friendship that felt this way? What was missing?
- ③ In the three-generation case study, how might technology have helped each generation maintain friendships? What virtues might have been harder to develop in each era?
- ④ If you had to choose: 200 online friends you interact with daily through screens, or 10 close friends you see in person weekly? What does your answer reveal about what you value in friendship?

Understanding Collective-Action Traps (Why You Can't Just Quit)

- A **collective-action problem** occurs when everyone would be better off if all took a certain action, but each individual is deterred from acting alone.
- Example: Fishermen would all benefit from limiting their catch to preserve fish populations, but any individual who limits their catch just loses money while others overfish.
- Social media creates the same trap: most students would prefer if *no one* used Instagram, but each feels they must use it because everyone else does.
- In one study, students said they'd need to be paid \$50 to quit Instagram—but would actually *pay* to have everyone quit together.



Friendship Virtue Lost: The Four Features of Real Interaction

- Haidt identifies four features that characterize healthy human interaction throughout history, all of which are diminished or eliminated online.
- **Embodyed:** Real-world interactions use hands, facial expressions, and body language—channels for which we have evolutionary programming that virtual interactions lack.
- **Synchronous:** Real-time interactions teach us conversational turn-taking, timing, and help us feel "in sync" with others in ways that asynchronous texts cannot.
- **One-to-one or one-to-few:** Small-group interactions allow depth and authenticity, while broadcasting to large audiences makes communication performative and anxiety-inducing.
- **High barrier communities:** When exit is difficult, people invest in relationships and repair conflicts; when blocking is easy, relationships become disposable.

The Impact on Virtue

Without these features, young people develop into adults who are "less comfortable and less skilled at interacting in person"—undermining virtues like empathy, patience, and loyalty.

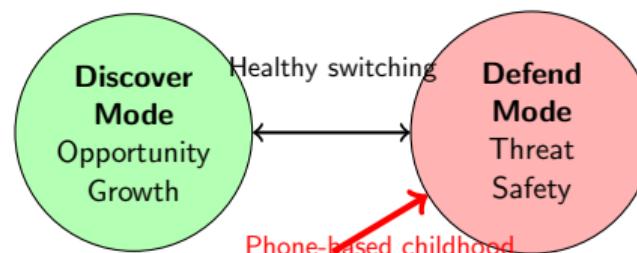
Courage and Risk: Why Children Need Dangerous Play

- **Play** is how young mammals wire their brains—practicing moves and skills they'll need as adults through vigorous, often risky activities.
- Children must take risks and fail in environments where failure has real but limited consequences—this is how they overcome fears and learn to estimate danger.
- **Thrilling play** (climbing, rough-housing, exploring unsupervised) appears most effective for overcoming childhood anxieties and building competence.
- When children are deprived of risk-taking opportunities, they develop into more anxious, risk-averse adults who struggle with challenges.

"Young people who are deprived of opportunities for risk taking and independent exploration will, on average, develop into more anxious and risk-averse adults." —Jonathan Haidt

Practical Wisdom: "Defend Mode" vs. "Discover Mode"

- Haidt explains that the brain has two systems: **approach** (when opportunities beckon) and **withdrawal** (when threats appear).
- People in "**discover mode**" see the world as full of opportunities—they are happier, less anxious, and more willing to engage with new ideas and people.
- People in "**defend mode**" see the world as full of threats—they are more prone to anxiety, depression, and hostile reactions to others.
- Phone-based childhood shifts young people's internal thermostats toward defend mode: constant social comparison, public shaming, and performance anxiety create chronic defensiveness.



Attention and Temperance: 237 Notifications Per Day

- **Temperance** is the virtue of moderation and self-control—particularly control over our desires and impulses.
- In 2023, the typical adolescent receives 237 notifications per day (roughly 15 every waking hour), each offering "high-pleasure, low-effort digital experiences."
- **Sustained attention** is essential for doing almost anything big, creative, or valuable—but notifications chop attention into fragments.
- When students have phone access during class, they use them (especially for texting and social media), and their grades and learning suffer measurably.

Activity	Requires Sustained Attention?	Compatible with 237 daily pings?
Deep reading	Yes	No
Creative writing	Yes	No
Learning math	Yes	No
Solving problems	Yes	No
Scrolling TikTok	No	Yes

The Loss of Civic Virtues: Echo Chambers and Spiral of Silence

- **Civic virtues** are character traits necessary for democratic participation: tolerance, willingness to compromise, ability to deliberate with those who disagree, and intellectual humility.
- Social media creates **echo chambers** (communities of like-minded people who reinforce each other's views) and **filter bubbles** (algorithmic curation that shows us only what we already agree with).
- The "Spiral of Silence" phenomenon intensifies online: people who hold minority or nuanced views stay silent for fear of being attacked by the mob.
- Both Vallor and Haidt argue that democracy requires shared reality and genuine dialogue—precisely what fragmented social media undermines.

The Democratic Crisis

When citizens can't agree on basic facts, can't listen to opposing views, and can't compromise, democracy itself becomes unsustainable. Social media accelerates this fragmentation.

Haidt's Four Norms for Breaking the Traps

- Because individuals can't escape collective-action traps alone, Haidt proposes four community-level norms that must be adopted together.
- These norms are designed to be implemented by parents, schools, and communities working in coordination—not by isolated families acting alone.
- Each norm addresses a different aspect of the phone-based childhood problem: devices, platforms, school environment, and real-world alternatives.
- Haidt argues that communities adopting all four norms will see substantial improvements in youth mental health within two years.

Haidt's Four Norms

- ① **No smartphones before high school** (age 14)—basic phones only
- ② **No social media before age 16**—delay account creation
- ③ **Phone-free schools**—locked pouches or lockers during school day
- ④ **More independence, free play, and responsibility**—in the real world

Case Study: Duckworth's National Survey—Early Findings on School Phone Bans

- Psychologist Angela Duckworth (University of Pennsylvania) is (in 2025-26) leading a national survey called "Phones in Focus" with over 20,000 public school educators responding.
- **Key finding:** "The stricter the policy, the happier the teacher. The stricter the policy, the less distraction there is for students in terms of their academic work."
- Teachers report dramatic changes after bell-to-bell bans: students make more eye contact, have more conversations at lunch (instead of "silently scrolling"), and are more engaged in class.
- However, only 1 in 4 high schools has a strict ban, compared to 3 in 4 elementary schools—meaning the students who most need support get the least.
- Teachers emphasize they face a **collective-action trap**: individual teachers can't enforce phone policies alone; they need school-wide and district-level support.

Discussion Questions: What Do Students Think About Phone Bans?

- ① Duckworth's survey asked teachers and principals, but not students yet. Based on your experience, do you think stricter phone policies would reduce distraction and improve learning? Why or why not?
- ② Teachers report that after phone bans, lunch periods became lively with conversation instead of silent scrolling. Would you see this as a benefit or a loss? What would you miss most about having your phone at lunch?
- ③ The research found that high schools have the most lenient policies even though teenagers face the strongest temptations. Do you think this is a mistake, or is there a good reason to trust high schoolers more than younger students?
- ④ Teachers said they can't enforce phone rules on their own—they need school-wide support. Have you seen examples of this collective-action trap in your school? What makes it hard for individual teachers to set rules?

Criticism Corner: Is Haidt Too Anti-Technology?

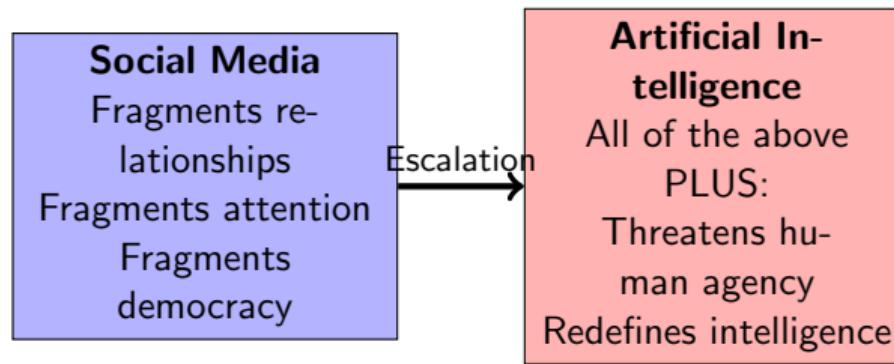
- **Correlation vs. Causation:** Critics note that Haidt shows correlation between smartphone adoption and mental health decline, but correlation doesn't prove causation—other factors (economic anxiety, academic pressure, COVID) might be responsible.
- **Overlooking Benefits:** Some researchers argue Haidt underemphasizes how social media helps isolated teens (LGBTQ+ youth, teens in rural areas, those with disabilities) find community and support.
- **Technological Determinism:** Critics claim Haidt treats technology as an unstoppable force that shapes us, ignoring how people actively choose how to use technology and can develop healthy habits.
- **Impractical Solutions:** Some argue his four norms are unrealistic—parents need to contact kids for safety, teens need phones for logistics, and complete bans might just drive usage underground.

Key Question

Even if Haidt's data shows correlation and his solutions have challenges, does that mean we should do nothing? What's the appropriate response to preliminary but concerning evidence?

Transition: From Social Media to Artificial Intelligence

- Both Haidt and Vallor see social media as undermining human flourishing by fragmenting attention, weakening relationships, and corroding civic virtues.
- But in recent years, a new technology has emerged that both authors view as potentially even more threatening: artificial intelligence.
- While social media changed how we interact with each other, AI threatens to change how we think, create, and understand ourselves.
- Vallor argues that AI presents many of the same dangers as social media (distraction, passivity, reduced human agency) but adds a new threat: the diminishment of what it means to be human.



Why AI Is Not a Mind: The Mirror Metaphor Explained

- When you look in a bathroom mirror, you see a face—but you know there isn't a second person behind the glass with thoughts, feelings, and experiences.
- Vallor argues that AI systems like ChatGPT work the same way: they produce **reflections of intelligence** without actually thinking, understanding, or experiencing anything.
- The AI was trained on billions of human texts and learned to produce patterns that *look like* human reasoning, creativity, and knowledge.
- Just as a mirror reflection has "no warmth, no depth," AI output lacks the experience, understanding, and meaning-making that characterize genuine thinking.

"When you go into the bathroom to brush your teeth, you know there isn't a second face looking back at you. That's just a reflection of a face, and it has very different properties. It doesn't have warmth; it doesn't have depth." —Shannon Vallor

The Dangerous Myth: "AI Thinks Like Us, But Better"

- Many AI researchers and tech leaders claim that AI systems are developing genuine intelligence and understanding—just like humans, but superior.
- This claim is dangerous not (just) because it overestimates AI, but because it **diminishes what human thinking actually is**.
- To make AI seem "intelligent like us," we must reduce human intelligence to mere pattern-matching, prediction, and statistical processing.
- Vallor argues this reductive view ignores the most important features of human minds: experience, consciousness, meaning-making, and genuine understanding.

What AI Actually Does	What Humans Actually Do
Pattern matching	Genuine understanding
Statistical prediction	Creative meaning-making
Probability distributions	Conscious experience
Optimization	Value-driven judgment

Geoffrey Hinton's Error: Confusing Performance with Understanding

- Geoffrey Hinton won the 2024 Nobel Prize in Physics for pioneering the deep-learning techniques that made modern AI possible.
- Yet Vallor argues he makes a fundamental philosophical error when he claims "we understand language in much the same way as these large language models."
- Hinton confuses the ability to produce correct outputs (performance) with genuine comprehension (understanding)—like mistaking a calculator's ability to produce "9" for $3+6$ with actual understanding of addition.
- Because AI can process vastly more data and respond faster than humans, Hinton concludes it might "supplant us" and that "humanity is just a passing phase in the evolution of intelligence."

The Philosophical Mistake

According to Vallor, Hinton treats **thinking** as mere information processing, erasing the role of consciousness, experience, and understanding—the very things that make thinking meaningful. What do you think?

AGI Redefined: From "General Intelligence" to "Replaces Workers"

- **Artificial General Intelligence (AGI)** originally meant a machine that could do everything a human mind can do—think, understand, feel, create meaning—at human level or better.
- Tech leaders like Sam Altman (CEO of OpenAI) have quietly redefined AGI to mean something much narrower: "a system equal to or better than humans at economically valuable tasks."
- Under this new definition, AGI doesn't need consciousness, understanding, or meaning-making—it just needs to do your job well enough that your boss can replace you.
- Vallor says this redefinition reveals Silicon Valley's values: "Now all we have for the target of AGI is something that your boss can replace you with. It can be as mindless as a toaster, as long as it can do your work."

Two Definitions of AGI

- **Original AGI:** A machine that thinks, understands, and experiences like humans (or better)
- **Redefined AGI:** A machine that performs economically valuable work (regardless of whether it "thinks")

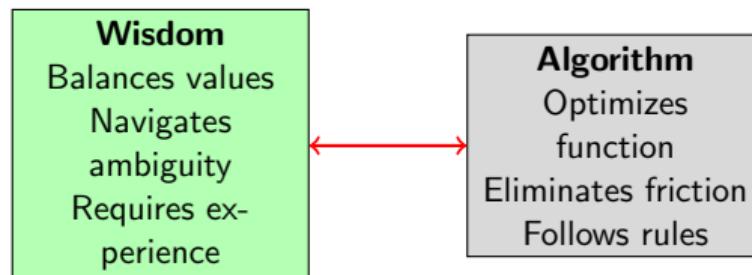
Silicon Valley's Religion: Efficiency as the Highest Good

- Vallor describes Silicon Valley's ideology as having "features of a religion"—certain commitments are "unshakeable by any kind of counterevidence or argument."
- The highest value in this worldview is **efficiency**: accomplishing tasks with maximum speed and minimum friction, regardless of any other considerations.
- But efficiency is meaningless without reference to some higher purpose. What is the point of being efficient if not to serve human flourishing?
- This ideology leads to viewing intelligence as "a thing that wants to remove the business of thinking"—optimization that eliminates difficulty, ambiguity, and the need for human judgment.

"It's about achieving a situation where the problem is solved and there's no more friction, no more ambiguity... you've dominated the problem and it's gone and all there is left is your perfect shining solution." —Shannon Vallor

The Virtue of Wisdom vs. Mechanical Problem-Solving

- **Practical wisdom** (or prudence) is the virtue of knowing what matters, what to do in complex situations, and how to live well—Aristotle considered it the master virtue.
- Vallor argues there is no "mathematical solution to the problem of justice" because justice involves "conflicting values and interests that cannot be made commensurable on a single scale."
- If we allow AI to handle increasingly complex decisions (hiring, medical treatment, criminal justice), we stop practicing the virtue of wisdom and lose confidence in human judgment.



Not all problems can or
should be algorithmic

Discussion Questions: Vallor on AI and Human Intelligence

- ① Vallor uses the mirror metaphor to explain why AI isn't really thinking. Do you find this metaphor convincing? Can you think of ways AI might be more than just a "reflection" of human intelligence?
- ② When Geoffrey Hinton says "we understand language in much the same way as large language models," Vallor says he's confusing performance with understanding. What's the difference? Can something perform perfectly without understanding?
- ③ Think about a time you truly understood something (a concept, a friend's feelings, a book). What was that experience like? Could a computer have that same experience by processing the right data?
- ④ Silicon Valley leaders redefined AGI from "thinks like humans" to "does economically valuable work." Why does Vallor think this redefinition is dangerous? Do you agree?

Criticism Corner: Does Vallor Underestimate AI's Significance?

- **The Existential Risk Camp (e.g., Geoffrey Hinton, Nick Bostrom):** Some AI researchers argue Vallor underestimates the possibility that AI could develop genuine consciousness or intelligence—and therefore pose existential threats. They worry her "it's just a mirror" view makes us complacent about real dangers.
- **The Techno-Optimists:** Some of these same researchers also believe AI could solve humanity's greatest problems—cure diseases, end poverty, achieve scientific breakthroughs—and that Vallor's focus on current limitations misses AI's transformative potential (for good or ill).
- **The Libertarian Critique:** Some argue Vallor's concerns about Silicon Valley and efficiency reflect a general skepticism toward capitalism and free markets, not specific problems with AI technology itself.
- **The "She'll Be Proven Wrong" Argument:** As AI capabilities advance rapidly, critics suggest her philosophical arguments about AI lacking understanding will seem outdated when systems demonstrate increasingly sophisticated reasoning.

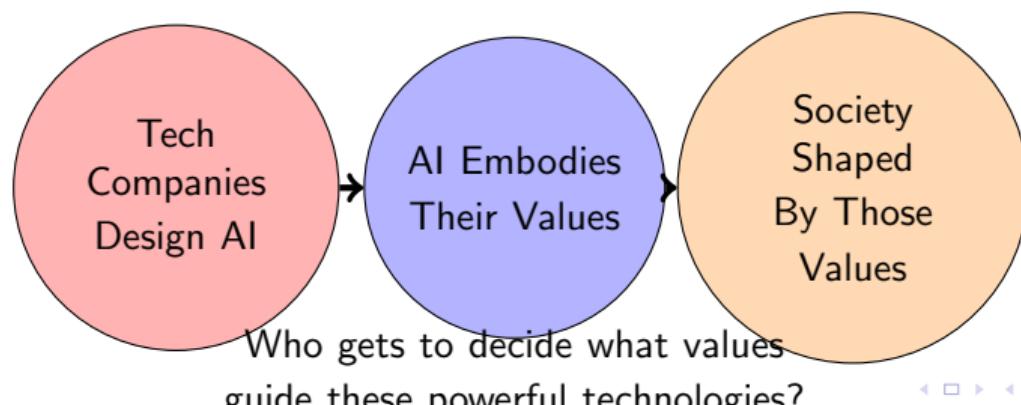
Vallor on Social Media and Democracy: The Fragmented Public Sphere

- Vallor's earlier work extensively examined how social media platforms undermine **democratic deliberation**—the process by which citizens reason together about shared problems.
- Social media creates **filter bubbles** where algorithms show us only content we already agree with, preventing exposure to diverse viewpoints necessary for democratic citizenship.
- The "many-to-many" communication of social media initially seemed liberating, but in practice it enabled mob dynamics, viral outrage, and the spread of misinformation at unprecedented scale.
- She argues that democracy requires not just free speech but **shared reality**—basic agreement on facts—which social media's fragmentation makes nearly impossible.

"When citizens lose trust in institutions, they lose trust in the stories told by those institutions... young people educated in the post-Babel era are less likely to arrive at a coherent story of who we are as a people." —Shannon Vallor

Both Authors on AI: Who Decides What Matters?

- Both Haidt and Vallor worry that AI systems are being designed and deployed by a narrow group of tech leaders with particular values and blind spots.
- Haidt notes that AI-powered recommendation algorithms on social media platforms already shape what billions see, read, and believe—concentrating enormous power in few hands.
- Vallor argues that if we allow AI to make increasingly important decisions (medical treatment, hiring, criminal justice, education), we're letting **Silicon Valley's values** (efficiency, optimization, quantification) become society's values.



Justice Online: Accountability, Doxing, and Cancel Culture

- Both authors grapple with questions of **justice** in online spaces: When is public shaming warranted? When does it become mob violence?
- **Doxing** (publishing someone's private information online) and coordinated harassment campaigns raise difficult questions about accountability versus cruelty.
- Some argue that social media finally enables justice by holding powerful people accountable through public exposure; others see dangerous vigilantism without due process.
- The virtue of justice requires both holding people accountable for wrongdoing AND protecting people from disproportionate or wrongful punishment.
- Haidt notes that social media makes it easy to "shoot darts" at others with no cost to ourselves, corroding the careful balance that justice requires.

The Justice Dilemma

How do we maintain accountability for harmful behavior while preserving fairness, proportionality, and the possibility of forgiveness and growth?

Intellectual Humility: Knowing What You Don't Know

- **Intellectual humility** is the virtue of recognizing the limits of your knowledge, being willing to revise your beliefs, and acknowledging when others might be right.
- Social media undermines this virtue by rewarding confident assertions and punishing admission of uncertainty or error—"I don't know" or "I was wrong" get few likes.
- AI systems present themselves with complete confidence even when they're wrong (what researchers call **hallucinations**), modeling a kind of arrogant certainty.
- In a world of infinite information and confident AI, the ability to say "I might be wrong" becomes more important, not less.

Intellectual Humility	Intellectual Arrogance
"I could be wrong"	"I'm definitely right"
Seeks contrary evidence	Ignores contradictions
Updates beliefs	Doubles down
Learns from others	Dismisses disagreement

Civic Friendship: Can We Talk Across Differences?

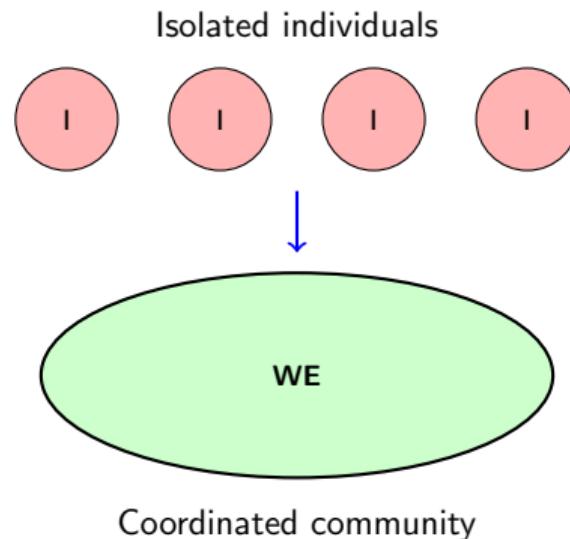
- **Civic friendship** doesn't mean being best friends with political opponents—it means treating fellow citizens as partners in shared democratic life, not enemies to be destroyed.
- Aristotle distinguished three types of friendship: utility (we help each other), pleasure (we enjoy each other), and virtue (we make each other better people).
- Civic friendship is closest to utility friendship: we may disagree deeply, but we need each other to make democracy work and to solve shared problems.
- Social media corrodes civic friendship by making it easy to dehumanize those we disagree with, to block them entirely, or to performatively attack them for social approval.

The Stakes

Climate change, pandemics, and economic challenges require cooperation across differences. Without civic friendship, we cannot solve collective problems.

The Virtue of Community: We Must Act Together

- Individual virtue is important, but both Haidt and Vallor emphasize that addressing technology's harms requires **collective action**—communities working together.
- Remember the collective-action trap: what's rational for each individual (staying on social media, using AI) may be harmful for everyone collectively.
- Historical examples: neighborhoods organizing to reduce crime, workers forming unions, citizens building public infrastructure—all required community virtue.



Hope: Technology Can Serve Human Flourishing

- Despite their criticisms, neither Haidt nor Vallor wants to eliminate technology or return to a pre-digital world—both see technology's potential for good.
- Vallor: "Technology at its core can be as humane an activity as anything can be. We've been technological creatures since before we were *Homo sapiens*."
- The question is not "technology or no technology" but "which technologies, designed how, used for what purposes?"

Virtue-Supporting Tech?

- Tools that enhance real-world connection (e.g., platforms for community organizing)
- AI that augments rather than replaces human judgment (e.g., decision-support systems)
- Technologies that promote civic engagement (e.g., tools for deliberative democracy)

Case Study: Democracy Under Pressure—The 2020 Election and Social Media

- Both Haidt and Vallor point to the 2020 U.S. presidential election as a case study in how social media can undermine democratic civic virtues.
- The QAnon conspiracy theory, spread primarily through Facebook, YouTube, and Twitter, convinced millions of Americans of elaborate false narratives—demonstrating social media's power to create alternative realities.
- After the election, false claims about voter fraud spread rapidly on social media despite being rejected by courts, election officials, and recounts—fragmenting shared reality.
- The January 6, 2021 Capitol insurrection was organized largely through social media platforms, showing how online radicalization translates to real-world violence.
- Haidt notes that social media made it "cheap and easy" for foreign actors (like Russia's Internet Research Agency) to "stoke rage on both the left and the right, often over race," further polarizing Americans.

Discussion Questions: What Civic Virtues Does Democracy Need?

- ① Both authors argue that democracy requires **shared reality**—basic agreement on facts. How can a democracy function when citizens live in completely different information universes? Is this situation sustainable?
- ② The 2020 election case shows how social media enables both grassroots organizing and violent extremism. Can we have the benefits (democratic participation) without the harms (radicalization and misinformation)?
- ③ Haidt emphasizes **civic friendship**—treating political opponents as fellow citizens, not enemies. Do you see examples of civic friendship online? What makes it so rare on social media platforms?
- ④ Vallor worries that social media discourse trains us to see every disagreement as a battle to win rather than an opportunity to learn. Have you noticed this in yourself or others? How might we cultivate **intellectual humility** online?
- ⑤ Both authors call for collective action—coordinated responses from communities, not just individual choices. What role should government play in regulating social media to protect democratic virtues?