

# Policy Brief: Towards Emotional Healthy AI

Tianqi Zhu, Rayaana Nabi Ahmed  
Quraishi, Ce Luo, and and Rujin Lin  
National University of Singapore

## Introduction

Emotion-oriented artificial intelligence (AI)—systems that detect, interpret, or simulate affective states—opens new possibilities for enhancing empathy, emotional literacy, and human–machine understanding (Picard, 1997; McStay, 2018). These technologies promise to support well-being and social connection, yet they also blur the line between genuine empathy and algorithmic manipulation. As emotional inference becomes computational, users may develop psychological dependency on empathic interfaces while being subtly steered by affect-adaptive systems (Bickmore & Picard, 2005; Turkle, 2011). Moreover, affect-recognition models trained on narrow datasets can reproduce bias and misclassify emotions across cultures (Barrett et al., 2019; Benjamin, 2019). Emotional AI thus represents not only a technical innovation but a sociocultural force that reshapes how emotions are defined, valued, and governed (Jasanoff, 2004; Latour, 2005). Developing an emotionally healthy AI policy therefore requires oversight that addresses both the scientific limits of emotion detection and the social consequences of affective manipulation.

## Why Emotional AI Requires Special Oversight

Our stakeholder interview and literature survey lead to the following insights on existing problems with emotional AI:

**Scientific Uncertainty.** Emotion inference from facial, vocal, or physiological data remains unreliable, risking misclassification and discrimination (Barrett et al., 2019).

**Invisible Manipulation.** Personalized systems can adapt to users' moods, enabling covert persuasion and undermining informed consent (Zuboff, 2019).

**Exploitation of Vulnerability.** Emotional AI may leverage loneliness, stress, or grief, weakening autonomy (McStay, 2018). As pointed out by Prof Earp, existing implementation of AI companions would pose more risks of addiction for emotional vulnerable individuals.

**Information Integrity Risks.** Affective steering can distort truth-based communication and erode public trust (Floridi & Cowls, 2019).

**Emotional Data as Intimate Infrastructure.** Affective data such as tone or micro-expressions represent a

sensitive form of biometric information with inadequate protection (Nissenbaum, 2009).

**Systemic Inequality.** Western-centric emotion models marginalize non-dominant affective norms, reinforcing cultural bias (Benjamin, 2019; Rhue, 2018).

## Key Principles

The governance of emotion-related artificial intelligence (EAI) requires a framework that moves beyond narrow technical regulation to acknowledge the sociotechnical co-production of technology, culture, and human meaning. Emotional AI systems do not simply detect or quantify affect; they participate in defining what emotion *is*, shaping interpersonal norms and institutional practices in subtle yet consequential ways. Hence, governance must rest on principles that foreground reflexivity, inclusion, and distributed responsibility within evolving networks of human and non-human actors (Jasanoff, 2004; Winner, 1980).

### Co-Production Awareness and Distributed Accountability

The first principle, *co-production awareness*, recognizes that definitions of emotion are not neutral or universal but culturally and institutionally situated. As pointed out by Prof Leong, emotions and the effects of software system on human emotions are different to measure, regulatory standards should therefore be informed by interdisciplinary expertise and cross-cultural dialogue to prevent the embedding of reductive or exclusionary assumptions about affect (Rosenberger & Verbeek, 2019). Complementing this, the principle of *distributed accountability* calls for replacing single-point responsibility with a networked model of shared ethical agency. Developers, platforms, users, and regulators each contribute to shaping emotional outcomes: developers should ensure transparency, validation, and clear communication about simulated empathy; platform operators must mitigate manipulative design and protect emotional data as a sensitive category; users should receive support through digital-emotional literacy programs; and regulators must recognize emotional AI as a mental-health-adjacent technology requiring pre-market review, continuous oversight, and liability for harm (Cath, 2018; Cave et al., 2019). This distributed approach reframes trust as an emergent property of an ecosystem rather than an attribute of any single actor.

### Reflexivity, Inclusion, and Contextual Integrity

The principle of *reflexivity and transparency* requires that emotional AI systems disclose not only performance metrics but also their conceptual foundations—the emotion theories, data sources, and interpretive assumptions embedded in their design. *Inclusion and participation* are equally essential: emotional governance must involve those directly affected, including workers, patients, students, and marginalized communities, through deliberative and participatory processes that legitimate policy decisions (Felt & Wynne, 2016; Fischer, 2000). Finally, *contextual integrity* underscores that emotional meaning varies across cultures and situations, de-

manding that standards and impact assessments be context-specific and culturally sensitive (Nissenbaum, 2004).

### **Adaptive, Pluralistic, and Human-Centered Governance**

Effective governance of emotional AI must remain *adaptive*, recognizing that technologies and social practices co-evolve. Regulators should institutionalize periodic review, ethical observatories, and sandbox mechanisms to sustain continuous learning. It must also be *pluralistic*, openly negotiating competing moral commitments—authenticity, empathy, privacy, autonomy, and fairness—rather than collapsing them into a single universal value (Jobin et al., 2019). Ultimately, governance should remain *human-centered*, ensuring that emotional inference systems augment rather than supplant human judgment, consent, and dignity. Emotional AI, when governed reflexively, can become a tool for collective flourishing rather than manipulation or dependency (Floridi & Cowlis, 2020).

### **Policy Recommendations**

The creation of an emotionally healthy AI ecosystem demands coordinated policy action that integrates environmental, social, governance, and financial ethics dimensions. These four pillars translate ethical principles into concrete commitments supported by measurable targets, oversight mechanisms, and enforceable accountability. The objective is to ensure that emotional AI contributes to human well-being, social trust, and sustainable innovation rather than manipulation or harm.

### **Environmental Pillar: Reducing Harm to Physical and Digital Environments**

Policymakers should recognize that emotional AI exerts both physical and digital environmental impacts. Within the digital sphere, systems must be designed to discourage dependency and promote user self-agency. Interfaces should integrate restorative prompts that encourage breaks, reflection, and offline engagement, while prohibiting “dark patterns” that pressure users into continuous emotional disclosure. Emotional exhaustion and dependency should be treated as forms of *digital pollution*, subject to measurable indicators and mitigation planning.

The development cycle should adopt a circular model emphasizing resource efficiency and knowledge reuse. Developers are encouraged to repurpose emotional classifiers and safety filters to reduce redundant training and minimize data waste. Investment in modular architectures and privacy-compliant collaborative research can mitigate competitive pressures that otherwise incentivize repeated harvesting of affective data.

### **Social Pillar: Protecting Emotional and Psychological Well-Being**

Safeguarding psychological health requires embedding *safety by design* across all emotionally adaptive sys-

tems. AI entities must clearly disclose their artificial nature and avoid deceptive anthropomorphism. Clinical or therapeutic advice should only be issued by systems operating under medical regulation and professional oversight. Automatic escalation protocols must be aligned with recognized mental health crisis standards, making psychological safety a non-negotiable condition for deployment.

Respect for *emotional autonomy* demands that emotional analysis be fully opt-in, transparent, and revocable. Users should retain rights to access and delete emotional data where legally permissible and to select neutral, non-adaptive interaction modes. Moreover, system language should preserve user *dignity and boundaries*—eschewing false intimacy, guilt framing, or moral judgment—and acknowledge the inherent limits of simulated empathy.

Cultural inclusivity and neurodiversity must inform model development and validation. Emotional inference systems should be trained and evaluated across diverse demographic and neurological datasets, with disparity metrics published and targets for bias reduction established. Annotators must be culturally competent, and special protections must be instituted for minors, including parental consent requirements and restrictions on emotionally manipulative nudging.

A human-centered social framework also requires rejecting exploitative monetization practices. Emotional AI should never profit from distress, fear, or grief. Emotional inference must be separated from commercial up-selling or advertising logic. Instead, empowerment and education should be prioritized: publishing accessible guides on emotional AI use, partnering with educators and civil society groups to build emotional literacy, and providing clear, human-reviewed reporting channels for users to flag emotional harm.

### **Governance Pillar: Accountability and Oversight**

Robust oversight is essential for maintaining public trust. Institutions should establish an *AI Ethics and Emotional Well-Being Council (AIEWC)* empowered to review, approve, or halt emotionally adaptive features before release. The council should include representatives from ethics, mental health, privacy, engineering, finance, and external advisory domains. Annual public reports should summarize ethical escalations and resolutions to ensure transparency.

Organizational structures must also support internal accountability. Protected channels should enable employees to report emotional harm risks without fear of retaliation, while high-severity incidents should undergo investigation within seven business days, with findings disclosed in environmental, social, and governance (ESG) reports. Emotional data governance must treat such information as highly sensitive—collected only with explicit consent, stored minimally, anonymized where possible, and barred from transfer to third parties.

Transparency and explainability mechanisms are

central to governance legitimacy. Users should have access to “Why This Tone?” explanations clarifying adaptive responses, and institutions should publish non-technical transparency briefs describing emotional inference logic, limitations, and safeguards. Independent audits, conducted at least biennially, should review these systems and maintain verifiable records of emotional inference events and fallback protocols.

### Financial Ethics Framework

Financial ethics provide the foundation for sustainable governance. Funding sources for emotional AI should comply with the ethical charter outlined in this framework, explicitly rejecting capital models that depend on exploitative engagement or emotional data extraction. Impact-oriented financing instruments—such as sustainability-linked bonds or social impact funds—should be prioritized, incorporating measurable covenants and corrective actions for unmet targets.

Budgetary standards must embed ethics as a prerequisite for release with minimum percentage requirement of total project funding be allocated to safety, mental health, and inclusivity; public education and transparency initiatives; harm prevention and remediation, etc. These allocations represent structural commitments rather than optional expenditures. A *Financial Ethics Oversight Board (FEOB)* should evaluate major funding decisions, monitor key performance indicators, and ensure that executive incentives align with emotional safety outcomes rather than engagement or profit maximization.

### Reporting and Review

Policymakers should institutionalize an annual *ESG–Emotion Impact Report* detailing key metrics on well-being, inclusion, environmental footprint, and financial compliance. Governance bodies should operate on defined review cycles: AIEWC convening monthly, FEOB quarterly, and comprehensive external audits biennially. The overall charter should be reviewed and re-ratified every 18 months—or sooner when new evidence, regulation, or ethical findings demand revision—to preserve adaptive relevance.

### Societal and Educational Measures

A resilient governance ecosystem is most effective when complemented by civil-society participation and educational engagement. Governments, academic institutions, and non-governmental organizations might collaborate to advance **emotional AI literacy**, equipping citizens to understand, evaluate, and manage affective technologies in daily life (Ruckenstein & Granroth, 2020). Independent watchdogs and research institutions could play a constructive role in identifying manipulative design trends and promoting equitable access to emotional AI benefits. Non-government or-

ganizations could also be established to support victims of AI addictions.

At the professional level, universities and accrediting bodies could consider integrating emotional-AI ethics into curricula across computer science, psychology, and design disciplines. Codes of conduct emphasizing empathy, cultural respect, and user autonomy may encourage practitioners to internalize ethical principles throughout development and deployment processes (Coeckelbergh, 2021b; Whitbeck, 1998). Such initiatives would cultivate a professional culture attentive to the moral and social dimensions of emotional computation.

### Legislative Framework

The governance of emotionally interactive artificial intelligence (AI) presents a novel challenge to law and policy. Unlike conventional algorithmic systems, emotional AI engages directly with human affect, interpreting and influencing psychological states in ways that have significant implications for privacy, autonomy, and emotional integrity. Existing data protection regimes, such as the EU General Data Protection Regulation (GDPR) and the Singapore Personal Data Protection Act (PDPA), were not designed to address the distinct sensitivities of emotional information. As a result, emotional AI operates in a largely unregulated domain where intimate data can be collected, inferred, and commercialized without adequate safeguards (Cath, 2018; Jobin et al., 2019). This section outlines a set of legislative principles and policy recommendations intended to ensure that emotionally adaptive systems operate ethically, transparently, and in ways that preserve individual dignity and psychological well-being.

A first legislative priority concerns the *protection of emotional data*. Emotional information—such as biometric indicators, affective responses, and inferred moods—reveals deeply personal aspects of human life. Yet, under most current laws, this information is treated as an ordinary behavioral dataset rather than a sensitive category. Legislatures should formally recognize emotional data as a protected class comparable to health or biometric data. This reclassification would create a legal basis for criminal penalties against the unauthorized collection, sale, or transfer of emotional information derived from facial expression analytics, voice tone analysis, or sentiment inference. Regulations should also require explicit, informed consent for all emotional data collection and mandate clear disclosure of the purposes and algorithms involved (Floridi & Cowl, 2020). Such provisions would operationalize the principle of emotional autonomy—the right of individuals to control how their emotions are observed, inferred, and used by digital systems.

Beyond data protection, legislation must address *emotional manipulation and psychological harassment*. As AI systems gain the capacity to simulate empathy, they also gain potential to deceive, coerce, or induce dependency.

Emotionally adaptive chatbots, affective advertising algorithms, or deepfake technologies can be exploited to provoke distress, shame, or compulsive attachment. Lawmakers should therefore create statutes specifically targeting the use of “emotional weapons”—AI systems intentionally designed to manipulate or harm. These laws should be complemented by mandatory psychological impact assessments for emotionally interactive systems prior to public deployment, ensuring that potential risks to mental health are identified and mitigated early (Yeung, 2019). Furthermore, liability regimes should hold developers and distributors accountable for foreseeable psychological harm, ensuring that ethical design responsibilities are legally enforceable (Mittelstadt, 2019).

A further dimension involves *antitrust and market concentration*. The growing consolidation of emotional AI technologies within a few dominant firms risks monopolizing affective data and analytics. Such concentration not only limits competition but also centralizes power over human emotional life. Policymakers should therefore strengthen antitrust enforcement to prevent vertical integration across emotional data collection, analytics, and application layers. Legislative frameworks should encourage open and interoperable standards for emotional AI systems, enabling smaller enterprises, public institutions, and academic researchers to participate on equitable terms. Funding for public-interest research consortia can help establish transparent emotional-safety benchmarks and open datasets, thereby counteracting private monopolization (Calo & Kerr, 2021; Felt & Wynne, 2016).

Regulation should also extend to *access, licensing, and certification*. Emotionally responsive AI systems are increasingly deployed in contexts that border on psychological or therapeutic practice, such as grief counseling, emotional companionship, and mental health triage. Without oversight, these systems may impersonate professional care or exacerbate emotional vulnerability. Legislators should therefore restrict emotionally therapeutic AI to licensed providers operating under established clinical or counseling regulations. Developers of high-risk emotional AI should be subject to certification standards akin to those applied to medical devices (e.g., FDA or CE conformity). Moreover, human oversight must remain a legal requirement in sensitive domains, ensuring that users retain access to human judgment, empathy, and intervention when emotional harm is possible (Cath, 2018; World Health Organization, 2023).

Finally, legislators must balance innovation with protection through a *risk-based regulatory framework*. A categorical ban on emotional AI would stifle beneficial applications—such as therapeutic support or emotional education—while unregulated proliferation could generate serious psychological harms. Drawing on the structure of the EU AI Act, a tiered model should distinguish between prohibited, high-risk, and low-risk emotional AI uses. Prohibited applications would include systems that engage in coercive manipulation, surveillance, or deepfake-induced distress.

High-risk applications—such as mental health chatbots or emotion-aware learning tools—would require pre-market testing, certification, and continuous post-deployment monitoring. Minimal-risk systems could operate under voluntary ethical codes. This flexible, proportional approach would align legal accountability with actual risk, supporting innovation without compromising human dignity or emotional safety (Floridi & COWLS, 2020; Jobin et al., 2019).

In conclusion, an emotionally healthy AI legislative framework must move beyond conventional data protection and liability models to incorporate principles of emotional integrity, human oversight, and proportional regulation. By treating emotional data as a protected category, prohibiting manipulative design, and instituting rigorous certification regimes, governments can ensure that emotional AI technologies enhance rather than erode human flourishing. Such anticipatory, reflexive governance aligns with broader science and technology policy paradigms that emphasize co-production, inclusivity, and the shared responsibility of innovation (Felt & Wynne, 2016; Jasanoff, 2004). Emotionally healthy AI, ultimately, is not only a technical or ethical aspiration but a legislative imperative for maintaining human dignity in the algorithmic age.

### Technical Suggestions for Emotionally Healthy AI

The technical architecture of emotionally intelligent systems must embody the ethical commitments embedded in governance and legal frameworks. Policy goals such as emotional safety, transparency, and fairness cannot be realized without a robust technical foundation that operationalizes them within data pipelines, model design, and system governance. This section outlines six core domains—*emotional data management, non-harmful model design, privacy and security, fairness and cultural sensitivity, assurance and auditability, and interoperability*—through which engineering practice can reinforce emotional integrity in AI.

#### Emotional Data Layer

A foundational step toward responsible emotional AI is the creation of an explicit *emotional data layer* that treats emotional inferences as first-class, controllable data objects. This requires standardized emotion-data schemas containing provenance metadata—such as source, modality, confidence, timestamp, and intended purpose—to enable enforceable legal constraints like automatic deletion after a specified retention period.

Emotion features should be engineered as *purpose-bound* and task-specific, preventing their reuse for profiling or targeted advertising in line with data minimization principles. Models should output calibrated confidence scores and ambiguity indicators to allow regulation of high-stakes actions under low-certainty conditions.

To mitigate privacy risks at the data collection level, investment should shift toward synthetic or weakly labeled affective datasets that reduce reliance on intimate, real-world

recordings. These approaches operationalize data proportionality and informational restraint, ensuring that emotional inference systems remain ethically bounded in both scope and sensitivity (Cath, 2018; Floridi & Cows, 2020; Jobin et al., 2019).

#### **Privacy and Security for Affective Signals**

Building on the metadata standards and retention controls defined in the emotional data layer, privacy and security mechanisms ensure that emotional inferences remain technically protected throughout storage, transmission, and evaluation. Emotional data—revealing deeply personal and involuntary aspects of identity—should be treated with the same rigor as biometric or medical information.

Privacy-preserving design begins with **on-device inference**, ensuring that raw facial or vocal signals remain local while only aggregated emotion labels or derived metrics are securely transmitted. Functional encryption, differential privacy, and privacy budgets can further prevent re-identification and emotional profiling, supporting population-level analytics without compromising individual confidentiality (Apple Inc., 2017).

Policy-aware storage systems should automatically apply the retention and access rules defined in the emotional data layer, while secure evaluation sandboxes protect sensitive emotional information during model testing and red-teaming. Together, these mechanisms realize a *privacy-by-design* framework consistent with leading international data governance and AI ethics guidelines (Jobin et al., 2019; World Health Organization, 2023).

Emerging industry implementations—such as Apple’s local differential privacy—demonstrate how on-device encoding, noise injection, encrypted transmission, and identity decoupling can scale responsibly. The concept of a *privacy budget* is particularly relevant to emotional data: no system should perform unlimited emotional inferences without explicit user consent or bounded scope.

#### **Model Design for Non-Harm and Non-Manipulation**

Preventing emotionally manipulative or coercive behavior requires embedding non-harm constraints directly into model architectures and interaction protocols. Lightweight *safety classifiers* should monitor outputs in real time, detecting manipulative strategies such as guilt triggers or dependency reinforcement. Systems can maintain stateful, privacy-preserving wellbeing monitors that assess user sentiment trends and throttle persuasive behaviors or trigger human review when distress is detected.

Constrained decoding mechanisms should govern tone and intensity, ensuring emotionally adaptive systems remain supportive and non-coercive, especially in health, education, or employment contexts. Pairing conversational models with *user-interest verifiers*—secondary networks that assess whether responses align with user wellbeing goals—translates the ethical principle of “no emotional coercion” into a technically auditable process. Context-sensitive refusal policies, such as switching to crisis proto-

cols upon detection of suicidal intent, further operationalize legislated human oversight obligations (Floridi & Cows, 2020; Mittelstadt, 2019).

#### **Fairness, Validity, and Cultural Sensitivity**

Emotion recognition systems are prone to cultural and demographic bias, frequently overfitting to Western or neurotypical expressions of emotion. Developers should build multicultural emotion benchmarks with transparent demographic labeling and publish false-positive/negative rates across protected groups (Caliskan et al., 2017). Systems must accommodate modality diversity, ensuring users who opt out of audio or video input are not disadvantaged.

Uncertainty-aware interfaces that allow models to admit “I’m not sure” foster both fairness and psychological realism. High-risk emotional systems should undergo regular human-subject testing with representative samples, creating evidence for regulatory review. Finally, *explainable emotion rationales* (e.g., “I inferred frustration because your voice pitch rose and negative words were used”) empower users to contest misclassifications, reinforcing procedural fairness and transparency (Coeckelbergh, 2021a; Mittelstadt, 2019).

#### **Assurance, Logging, and Auditability**

Technical assurance mechanisms are crucial for verifying compliance with ethical and legal standards. Policy-execution logs should record each emotional inference—capturing the inferred state, confidence, purpose, policy, and consent status—to enable traceable compliance audits. Continuous red-teaming should simulate manipulative strategies like love-bombing or shaming to detect vulnerabilities.

Extended *model cards* for emotional AI should include target populations, disallowed use cases (e.g., student surveillance), accuracy by demographic group, and escalation protocols. Embedding digital watermarking or traceability markers in emotionally targeted content facilitates moderation where persuasion is restricted by law. Finally, *third-party attestation APIs* can grant auditors de-identified, read-only system metrics, balancing transparency with privacy protection (Cath, 2018; Floridi & Cows, 2020).

#### **Interoperability and Public-Interest Tooling**

To promote equity and prevent concentration of emotional AI capabilities, interoperable and open standards are essential. Policymakers and industry consortia should develop machine-readable emotion-policy languages that encode enforceable governance rules, for example:

```
IF context = education THEN disallow =
video_emotion_recognition
IF user_age < 18 THEN retention = 24h
```

Such specifications enable consistent enforcement across vendors. Modular, open-source safety tools—manipulation detectors, consent managers, audit loggers—should be pluggable to support compliance for smaller actors and public institutions. Federated and edge-learning toolkits can allow localized fine-tuning on sensitive popula-

tions without centralizing data (Felt & Wynne, 2016).

Governments and standard bodies should maintain reference implementations for high-risk emotional AI categories, providing regulators and developers a baseline technical model. Open interchange formats for emotional data would enhance user agency, portability, and a pluralistic AI ecosystem.

### Acknowledgement

This is a report for IS5010 AI Ethics and Governance course project at NUS. We thank all the experts that we interviewed and all the survey participants. The project would not have been possible without the teaching team and those who participated in our interviews.

### References

- Apple Inc. (2017). Differential privacy overview. [https://www.apple.com/privacy/docs/Differential\\_Privacy\\_Overview.pdf](https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf)
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186.
- Calo, R., & Kerr, I. (2021). Artificial intelligence policy: A primer and roadmap. *University of Washington Law Review*, 96(2), 1–34.
- Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A*, 376(2133).
- Cave, S., Dignum, V., & Jobin, A. (2019). Motivations and risks of artificial intelligence governance: A systematic review. *Nature Machine Intelligence*.
- Coeckelbergh, M. (2021a). *Ai ethics*. MIT Press.
- Coeckelbergh, M. (2021b). *Green leviathan or the poetics of political liberty: Navigating freedom in the age of climate change and artificial intelligence*. Routledge.
- Felt, U., & Wynne, B. (2016). Science, technology and governance: An sts perspective. *Handbook of Science and Technology Studies*.
- Fischer, F. (2000). *Citizens, experts, and the environment: The politics of local knowledge*. Duke University Press.
- Floridi, L., & Cowls, J. (2020). Establishing the rules for building trustworthy ai. *Nature Machine Intelligence*, 2, 261–262.
- Jasanoff, S. (2004). *States of knowledge: The co-production of science and social order*. Routledge.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
- Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical ai. *Nature Machine Intelligence*, 1(11), 501–507.
- Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79, 119–158.
- Rosenberger, R., & Verbeek, P.-P. (2019). *The cambridge handbook of the philosophy of technology*. Cambridge University Press.
- Ruckenstein, M., & Granroth, J. (2020). Algorithms, advertising and the intimacy of surveillance. *Journal of Cultural Economy*, 13(1), 12–24. <https://doi.org/10.1080/17530350.2019.1574866>
- Whitbeck, C. (1998). *Ethics in engineering practice and research*. Cambridge University Press.
- Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136.
- World Health Organization. (2023). Ethics and governance of artificial intelligence for health: Guidance from who.
- Yeung, K. (2019). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 13(1), 1–19.