Question (Title):

How should whistleblowing frameworks adapt when tech companies' informational harms affect billions but lack the physical impact of traditional engineering failures

You're a senior cybersecurity engineer who discovers your company's algorithms are amplifying dangerous misinformation, and executives have known for months but won't fix it because it would reduce ad revenue. Your manager orders you to stay silent. You're four months from stock vesting, your team depends on you, but you're also bound by the ACM Code of Ethics requiring you to "contribute to society and to human well-being" (ACM, 2018). This is the "divided loyalties" problem from our lecture materials, where professional obligations to multiple stakeholders come into fundamental conflict.

De George (1999) says whistleblowing is **permissible** when: (1) harm is serious, (2) you've reported internally, (3) you've exhausted all channels, (4) you have documented evidence, and (5) going public will likely prevent harm. It becomes **obligatory** only when you meet the first three conditions PLUS have strong evidence of effectiveness. But here's the tension: De George's framework was built for traditional engineering contexts involving physical safety, bridges collapsing, medical devices malfunctioning, airplane systems failing. When Frances Haugen disclosed Facebook's internal research showing Instagram's documented harm to teenage girls' mental health in 2021, she had documentation, exhausted internal channels, and revealed harm at massive scale, but the harm was psychological and social, not physical (Horwitz & Seetharaman, 2021; U.S. Senate, 2021).

Additionally, Nissenbaum (1994) describes the "many hands" problem: in tech companies, thousands of people contribute to harmful outcomes through complex systems where responsibility becomes diffused across engineers, product managers, executives, and algorithms. If you're one of the few people who can see the full picture of how these distributed decisions create societal harm, does that specialized knowledge create a heightened individual obligation to act? Or does organizational complexity actually distribute responsibility in ways that reduce what any single professional should be expected to risk? The lecture materials suggest loyalty should extend to "society as a whole,

especially where safety and health issues are at stake," but they don't fully resolve whether informational harm at billion-user scale creates the same professional duties as physical harm in traditional engineering.

What I'm asking the community to discuss: In what specific ways would you modify De George's five criteria to address the unique characteristics of tech industry whistleblowing, particularly the scale of impact, the nature of informational versus physical harm, and the "many hands" problem? When applying the ethical frameworks we studied (Kantian deontology, utilitarian consequences, Rawlsian justice), how should we weigh an engineer's specialized knowledge against organizational complexity when determining individual moral obligation? What would a modernized whistleblowing framework look like that accounts for platform technologies affecting billions while still protecting professionals from being required to be "moral heroes" as Ladd (1991) warned against?

Sources:

ACM Code 2018 Task Force. (2018). *ACM Code of Ethics and Professional Conduct*. https://www.acm.org/code-of-ethics

De George, R. T. (1999). *Business ethics* (5th ed.). Prentice Hall.

Horwitz, J., & Seetharaman, D. (2021, September 14). Facebook knows Instagram is toxic for teen girls. *The Wall Street Journal*. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739