

Discussion 2: Establishing an Ethical Baseline for AI Development

Discussion Topic:

1. Context

There has been a rise in the development of artificial intelligence (AI) as a technology. There are various applications of AI and there are also numerous stakeholders that are involved. Ethical considerations must be made to ensure that these technologies are safe for human exposure.

2. Discuss

In this discussion, address the following:

- Address this focal problem by suggesting elements needed for producing a baseline for ethics in AI-based software development.
- Support your statements by citing cases of recent adaptations of AI for a particular academic, government, or industrial sector.

My Post:

Hello class,

Humanity is experiencing an AI revolution. A technological revolution that will impact humanity in unprecedented ways, in comparison, the industrial revolution's impact on society was just child's play... Just a year ago, in my Bachelor of Science computer science ethics class, in a class discussion, I remember arguing with classmates about the potential impact that AI will have on software development, such as replacing junior software engineers for coding tasks, which was, for most of my classmates, not feasible. Today, a year later, 90% of the code written by Anthropic teams is done by AI (Weinberger, 2025). In essence, Claude Code is writing its own code. To say the least, this is having an immense impact on how coding is done. We are just seeing the tip of the iceberg... we have just begun to realize how AI software could transform society. Consequently, as AI is being embedded in critical infrastructure and daily life, In essence, Claude Code is writing its own code. To say the least, this is having an immense impact on how coding is done. We are just seeing the tip of the iceberg... we have just begun to realize how AI software could transform society. Consequently, as AI is being embedded in critical infrastructure and daily life, establishing a robust ethical baseline is no longer optional.

Building an ethical baseline in AI-based software development requires moving away from just having abstract ethical principles (episteme approach) to an approach that not only formulates ethical principles but also actively applies these principles in practice. Additionally, this baseline should include a taxonomy mapping principles that include actionable and enforceable rules, as well as measurable outcomes of these rule implementations:

Principles → Policies → (Standards / Controls / Processes) → Measurable outcomes (O'Keefe & Brien, 2023)

The AI ethical baseline should include:

1. A governance and accountability element defining “who owns the harm?”
2. A risk classification and impact assessment element (before building, not after harm).

3. A data governance element covering the provenance, consent, privacy, and representativeness of AI training datasets and data queries.
4. An evaluation and control gates element assessing bias, robustness, security, and misuse testing.
5. A transparency, documentation, and contestability protocols element.

Recently, the academic, government, and industrial sectors started implementing various aspects of this ethical baseline. For example:

- In the academic sector, Arizona State University, in collaboration with OpenAI, implemented a campus-wide GenAI usage for teaching, research, and administrative functions that is focused on prioritizing equity, privacy, and human-centered design, ensuring that our AI solutions deliver social benefit while respecting individual rights and dignity. (Salcido, 2024; ASU, n.d.)
- In the government sector, the Office of Management and Budget issued M-25-21 (2025), the memorandum replaces earlier AI guidance and sets new expectations for governance, innovation, and public trust in federal AI use that:
 - remove barriers to innovation and provide the best value for the taxpayer
 - empower AI leaders to accelerate responsible AI adoption
 - ensure the federal agencies' use of AI works for the American people
- In the industrial sector, Microsoft implemented Microsoft's Responsible AI Standard v2 internal (Microsoft Corporation, 2022), representing an internal requirements-style approach (i.e., enforceable controls rather than broad ideals) for developing and deploying AI systems.

-Alex

References:

- ASU (n.d.). *Artificial Intelligence*. Arizona State University. <https://ai.asu.edu/>
- Microsoft Corporation. (2022). *Microsoft responsible AI standard v2: General requirements* [PDF]. <https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft-Responsible-AI-Standard-General-Requirements.pdf?country=us&culture=en-us>
- O'Keefe, K., & Brien, D. O. (2023). Chapter 03: Ethical principles, standards and practice. *Data ethics: Practical strategies for implementing ethical information management and governance*. Kogan Page.
- Office of Management and Budget. (2025, April 3). *M-25-21: Accelerating federal use of AI through innovation, governance, and public trust* (Memorandum) [PDF]. The White House. <https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf>
- Salcido, N. (2024, January 18). *Arizona State University collaboration with OpenAI charts the future of AI in higher education*. ASU News. <https://newsroom.asu.edu/press-releases/arizona-state-university-collaboration-openai-charts-future-ai-higher-education?>

Weinberger, M. (2025, October 16). *Most Anthropic teams are coding with AI, but it's not replacing humans*. *Business Insider*. <https://www.businessinsider.com/most-anthropic-teams-coding-with-claude-ai-not-replacing-humans-2025-10>