A Practical Framework to Evaluating Ethical Implications of Artificial Intelligence (AI) Systems

This framework provides questions and factors to consider when assessing ethical considerations of AI systems.

1. Micro-Lens Thinking: Specific Questions

Foundational Values

- Whose foundational values shape the system — users, vendors, or others?
- How are these embedded in the design?

Discipline-Specific Ethics

- What ethical standards apply to the field?
- How are they reflected in the AI system?

Data & Trust

- Is the system built with supervised learning or unsupervised learning?
- Are the training datasets trustworthy, with evidence to confirm this?
- Does the system meet copyright and procurement standards?
- What are the limitations of the system and how may they impact outcomes?

Problem Definition

- What problem is being solved?
- Holistically thinking, is AI the best solution?
- Who benefits from using AI and is there a net community gain?
- What problems may be created by using the system?

Stakeholders

- Who are the stakeholders?
- How is their privacy protected?
- What are the short, medium, and long-term impacts?
- Are these impacts justified?

Version 1.1 Date: 30/09/2025

Inclusion & Fairness

- How are special needs identified and addressed?

- What biases exist and what impact may they have?

- Is there built-in assurance against discrimination?

Transparency & Accountability

- How will stakeholders be informed?

- Where does accountability lie and is there any responsibility gap?

- What human oversight is in place?

- Is there a right of review process?

Evaluation & Maintenance

- What does success look like from a quality perspective?

- Is there an evaluation loop that revisits these questions?

- How is maintenance approached and will it have human oversight?

- What triggers a pause or shutdown?

- Can full and documented transparency be provided?

2. Macro-Lens Thinking: Factors to Consider

Building on the micro level thinking considered thus far, a macro level review against ethical theories and system lifecycle helps one to conclude if the means justify the end.

Because AI can be applied does not mean that AI *should* be applied.

To guide whether you *should*, explore these ethical theories and integrate into your analysis one to two theories which best fit the context:

Ethical Theory	Key Question	Focus
Utilitarianism	Will this lead to the greatest good for the greatest number?	Net positive outcomes
Deontology	Do we have a duty to act, regardless of consequences?	Moral obligation

Version 1.1 Date: 30/09/2025

Ethical Theory	Key Question	Focus
Virtue Ethics	What would a virtuous person do?	Courage, honesty, compassion, amongst others
Care Ethics	How does this affect vulnerable people and the environment?	Relationships and care
Principilism Ethics	Does it avoid harm, promote good, respect autonomy, and ensure fairness?	Holistic
Indigenous Peoples' ethics	Does it impact connection to land, community, and ancestral traditions?	Harmony with nature, collective well-being, and intergenerational stewardship

Lifecycle: Does the Means Justify the End?

Evaluate the full lifecycle (i.e. training datasets &system development onwards) of the AI system based on the following matters:

- **Intellectual Property & Plagiarism**

- **Environmental Impact**

- **Bias & Fairness**

- **Privacy & Data Security**

- **Misinformation Risks**

- **Socioeconomic Impact**

- **Educational Impact**

- **Human Health & Growth**

- **Autonomous Decision-Making**

- **Trustworthiness**

Version 1.1 Date: 30/09/2025