

AI Systems – For the People, Of the People, By the People

As human beings, we have never fully deciphered the mysteries of our own brain. Understanding how it truly functions remains an ongoing and incomplete journey. Yet, paradoxically, we have embarked on the ambitious task of creating a system we call *Artificial Intelligence*, an attempt to mimic the very organ we do not yet fully understand. It is a striking irony: we strive to replicate what still eludes our own comprehension.

To add to the intrigue and complexity, this journey is unfolding at a breathtaking pace. We are attempting to solve human problems with a machine that must itself adhere to human principles — governed by human standards, designed to be compliant, and expected to be risk-averse.

The cycle itself highlights the paradox:

- We create a **business case** to address a human problem.
- We use **AI** to design a solution.
- We then apply **risk management**, not only to manage human risks, but also to manage *machine risks*, which are now intertwined with human risks. **A double burden.**

In this pursuit, we demand that the machine remain compliant, even as humans create, manage, monitor, and operate it.

And above all, we must ensure the machine poses no harm: not to human beings, not from human beings, and not by human beings.

Table below is a thoughtful attempt to draw parallels between human qualities and AI system characteristics

Human Being Qualities	Explanation	Artificial Intelligence System Characteristics	Explanation	Principle(s) Supported	Principle Explanation
Brain	Central processing organ; adaptive, self-organizing, rewires through neuroplasticity.	Algorithm + Model Architecture	Rules + structural design (e.g., neural networks) enabling AI to process and learn.	Reliability	Reliable foundations come from strong architecture.
Experience	Raw lived events, unstructured and accumulated.	Data	Raw inputs collected from the world, unstructured until processed.	Reliable Outcomes	Consistency requires a solid base of diverse data.
Knowledge	Structured, internalized understanding built from experience.	Training	The process that structures raw data into a model's internalized "knowledge".	Reliability, Accuracy	Accuracy depends on correct, unbiased training.
Wisdom	Applying knowledge with foresight, ethics, and judgment.	Contextual Optimization & Feature Engineering	Selecting and refining what matters from data to enable effective decision-making.	Ethics	Ensures AI decisions reflect values and judgment.
Due Care	Acting responsibly and with caution.	Guardrails / Governance	Controls and constraints ensuring AI	Ethics, Safety, Prevention of Harm	Prevents unintended

Human Being Qualities	Explanation	Artificial Intelligence System Characteristics	Explanation	Principle(s) Supported	Principle Explanation
			operates within safe and ethical boundaries.		harm by setting boundaries.
Due Diligence	Thorough investigation and validation before acting.	Testing & Validation	Evaluation of models to ensure reliability, accuracy, and safety before deployment.	Safety, Reliability, Accuracy	Ensures robust and precise outcomes.
Communication / Reasoning	Ability to explain and justify decisions.	Explainability	AI's ability to communicate decision logic in human-understandable terms.	Explainability, Transparency	Builds trust through clear reasoning.
Trust	Confidence built on consistent and transparent behavior.	Interpretability & Transparency	Mechanisms that make AI decisions understandable and trustworthy.	Transparency, Reliable Outcomes	Users trust systems they can see into.
Bias	Cognitive distortion or prejudice influencing judgment.	Bias (systemic skew) + Detection & Mitigation	AI bias from data/algorithms, managed via monitoring and correction.	Ethics, Fairness, Prevention of Harm	Prevents discriminatory or harmful outcomes.
Accountability	Being answerable for actions and decisions.	Human Oversight, Governance & Liability	Humans remain responsible for AI actions, ensuring governance and liability.	Ethics, Transparency	Keeps responsibility with humans, not machines.
Fairness	Equitable and just treatment of all.	Fairness	Ensuring AI outcomes are unbiased, equitable, and aligned with ethical principles.	Ethics, Reliability, Prevention of Harm	Protects rights, ensures equitable treatment.

Summary mapping of the Principle

- **Ethics** - Aligns AI with values (Wisdom, Due Care, Fairness, Accountability).
- **Transparency** -Users must understand decisions (Communication, Trust, Accountability).
- **Safety** - Prevents unsafe use (Due Care, Due Diligence).
- **Reliable Outcomes** - Ensures consistency (Brain, Experience, Knowledge, Trust, Fairness).
- **Explainability** - Enables human understanding (Communication / Reasoning).
- **Accuracy** - Ensures precision & correctness (Knowledge, Due Diligence).
- **Prevention of Harm** - Avoids negative impacts (Due Care, Bias, Fairness).

Author : Roshan Neville Sequeira

Date : 17th August 2025

Place – Abu Dhabi, United Arab Emirates