The Human Imperative in Artificial Intelligence

Hunjun Shin

1 What is Human-Centered AI?

Human-Centered Artificial Intelligence (HCAI) is an approach to designing AI systems that places people at its core, aiming to create technologies that amplify, augment, enhance, and empower humans, rather than simply replacing them. It differs from purely technology-centric approaches by prioritizing human well-being and control over raw technical capability.

1.1 The Core Principles of HCAI

- **Two-Dimensional Framework:** A central principle of HCAI is to move beyond the traditional one-dimensional view where more automation means less human control. HCAI advocates for a two-dimensional framework that seeks high levels of human control and high levels of computer automation simultaneously. This approach supports human self-efficacy, mastery, and creativity.
- **Reliable, Safe, and Trustworthy (RST) Systems:** HCAI aims to produce AI systems that are inherently RST.
  - **Reliability** refers to consistent, predictable technical performance.
  - **Safety** focuses on preventing harm to people or the environment.
  - **Trustworthiness** involves external validation, ensuring the system operates according to established ethical norms and societal values.
- **Broad Stakeholder Consideration:** All parties affected by an AI system, not just direct users, should be considered during the design process to ensure broadly accepted and equitable solutions.
- **Ethical Operations:** HCAI systems prioritize key ethical attributes:
  - **Transparency:** Users should understand what the AI system is doing and why.
  - **Fairness:** Systems must be developed to help all people equitably, critically examining data and metrics to eliminate bias.
  - **Accountability:** Systems should be accountable, often through independent oversight structures that ensure responsible operation.

2 Why HCAI Matters

HCAI is critical because it shifts the focus of AI development from purely technical capabilities to human well-being, ensuring systems are effective, ethical, safe, and empowering.

2.1 Ethical Considerations

A technology-centric approach often overlooks profound ethical implications. For example, the **COMPAS recidivism prediction system** was technically "well calibrated," yet ProPublica found it produced false positives that harmed Black defendants twice as often as White defendants. This shows that maximizing a technical metric does not guarantee fairness; fairness is a "societal values question" requiring human judgment. HCAI also stresses accountability through independent oversight to prevent "algorithmic hubris" from leading to harmful autonomous systems.

2.2 Ensuring Safety and Reliability

HCAI's focus on RST systems is vital, as failures without human consideration can be catastrophic. The **Boeing 737 MAX crashes** exemplify a "failure from excessive automation," where pilots were not trained to override an autonomous system. Similarly, the name **Tesla Autopilot** may encourage driver over-reliance, despite requiring active supervision. HCAI promotes safety through transparent management and technical practices like audit trails, while trustworthiness is built via independent oversight.

2.3 Augmenting Human Skills

HCAI's philosophy is to empower people, not replace them. The question is not "Will AI replace radiologists?" but rather "Radiologists who use AI will replace radiologists who don't." This is achieved through the two-dimensional framework, which allows high automation and high human control to coexist. A modern digital camera, for instance, offers extensive automatic adjustments (high automation) while still giving users granular creative control (high human control).

3 Connect Theory to Practice: Patient Controlled Analgesia

In HCAI, achieving the right **balance** between human control and computer automation is paramount. This balance is critical because both extremes are dangerous: excessive human control can lead to mistakes, while excessive computer control (as seen in the Boeing 737 MAX) can lead to a lack of human understanding and intervention. The optimal balance involves designing systems where automation handles routine or complex calculations, while humans retain meaningful oversight, decision-making power, and the ability to adapt.

This approach directly challenges the traditional, “mind-limiting” view of automation (e.g., SAE levels for self-driving cars) which suggests that human control must necessarily decrease as automation increases. A Patient Controlled Analgesia (PCA) device is a prime example of HCAI’s balanced, two-dimensional framework in practice.

A HCAI-designed PCA device empowers the patient to control their own pain medication, supporting self-efficacy. This high degree of human control is balanced with high automation for safety: the system uses interlocks to prevent overdosing. Furthermore, the system is made Reliable, Safe, and Trustworthy (RST) through:

- **Transparency:** The device provides explanations to the patient about its operation and the importance of dose limits, building trust.
- **Accountability:** A hospital control center monitors hundreds of PCA devices, reviewing audit trails and collecting data to improve future versions. This addresses the broader community of stakeholders (staff, other patients) and societal goals for safe healthcare.

By balancing patient control with automated safety features and providing oversight, the PCA device becomes a powerful example of a system that is both highly functional and deeply human-centered.

4 Future Challenges

The greatest future challenge for AI is not technical but social: defining the “right objective”. As the COMPAS case showed, what is technically optimal may be socially unjust. Deciding on metrics that represent our collective values is a “societal values question” that only humans can answer.

HCAI’s mandate to consider diverse stakeholders—users, communities, society—means reconciling conflicting needs and perspectives, a task requiring human arbitration. AI can optimize for predefined metrics, but it cannot understand or adapt to the nuanced, evolving definitions of “good” or “fair”.

Therefore, humans must remain the ultimate arbiters, providing essential ethical oversight and value judgments. HCAI’s principles of transparency and the two-dimensional framework are designed to keep humans in the loop, ensuring they retain mastery and responsibility. The challenge lies in navigating diverse human values, which inherently requires that humans guide the development, deployment, and governance of AI.