ISSN Online : 2709-5088
ISSN Print : 2709-507X
IJCST

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY (IJCST) Vol. 3 No. 1 (2019)

Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy

1st Ahmad Iqbal
HCI Lab,Information Technology University Lahore, PK ahmad.iqbal@itu.edu.pk
2nd Arfan Jaffar
Dept Computer Science Superior University, Lahore Lahore, PK arfan.jaffar@superior.edu.pk

***Abstract*:**
In the realm of technology, well-crafted designs that strike a balance between human control and computer automa- tion can significantly enhance human performance, leading to widespread acceptance and adoption. The Human-Centered Ar- tificial Intelligence (HCAI) framework provides clear guidelines for achieving this delicate balance: (1) designing interfaces that optimize both human control and computer automation to en- hance human performance, (2) discerning situations where either full human control or complete computer control is essential, andsteering clear of the pitfalls associated with excessive human intervention or overreliance on computer algorithms. Employing the principles of HCAI enhances the likelihood of creating designs that are Reliable, Safe, and Trustworthy (RST). Accomplishing these objectives holds the potential to significantly boost human performance, empowering individuals while nurturing their self- confidence, expertise, creativity, and sense of responsibility.

***Keywords*:** BCI, Artificial Intelligence, Virtual Reality

INTRODUCTION

This paper presents a groundbreaking perspective on Human-Centered Artificial Intelligence (HCAI) by introduc- ing a novel framework that separates levels of automa- tion/autonomy from levels of human control. Unlike traditional approaches, the proposed guideline advocates for the simulta- neous pursuit of high levels of human control and high levels of automation, aiming to create computer applications that are Reliable, Safe, and Trustworthy (RST). This shift in focus offers a promising avenue, particularly for addressing com- plex and poorly understood problems, leading to significant improvements in human performance. Moreover, it supports human self-efficacy, mastery, creativity, and responsibility.

Venigandla, K., & Tatikonda, V. M. (2021) explain Diagnostic imaging analysis plays a pivotal role in modern healthcare, facilitating the accurate detection and characterization of various medical conditions. However, the increasing volume of imaging data coupled with the shortage of radiologists presents significant challenges for healthcare systems worldwide. In response, this research paper explores the integration of Robotic Process Automation (RPA) and Deep Learning technologies to enhance diagnostic imaging analysis.

Historically, the field of artificial intelligence has been influenced by the belief in computer autonomy, leading to the development of various levels of automation/autonomy. However, this paper challenges the conventional wisdom by emphasizing the importance of human control alongside automation [1]. The traditional one-dimensional perspective, where increased automation often comes at the expense of reduced human control, has limitations. Critics have pointed out the flaws in such an approach, highlighting issues related to human effort in monitoring autonomous systems and the resultant inferior performance [2]. Recognizing these chal- lenges, leading artificial intelligence researchers and devel- opers are increasingly acknowledging the need for human- centered designs[3]. This paradigm shift is essential to ensure the integration of creativity support features in technology, fostering innovation and enhancing user experience.

I. RELIABLE, SAFE & TRUSTWORTHY SYSTEMS: A HOLISTIC PERSPECTIVE

In this section, the discussion centers around the critical components that contribute to achieving Reliable, Safe, and Trustworthy (RST) systems. The definitions provided em- phasize the significance of technical practices, management strategies, and independent oversight structures in ensuring high performance, user confidence, and system integrity.

A. Technical Practices for Reliability

ISSN Online : 2709-5088
ISSN Print : 2709-507X
IJCST

Reliable systems, as defined by Modarres et al. (2016), are fostered through appropriate technical practices that support human responsibility, fairness, and explainability [4]. Key practices include:

- Implementing audit trails and analysis tools to review failures and near misses.
- Conducting benchmark tests for verification and valida- tion purposes.
- Continuous review of data quality and bias testing to adapt to shifting contexts.
- Designing interfaces that inspire confidence across di- verse user groups.
- Employing explainable user interfaces to minimize the need for detailed explanations

B. *Cultivating Cultures of Safety*

Cultures of safety are nurtured through open management strategies, encompassing:

- Demonstrating leadership commitment to safety within organizational hierarchies.
- Encouraging open reporting of failures and near misses.
- Establishing internal oversight boards to address prob- lems and plan for the future.
- Providing transparent public reports outlining problems and future plans.

C. *Independent Oversight Structures for Trustworthiness*

To instill trust, it is imperative to establish independent oversight structures [6]. These include:

- Professional organizations (e.g., IEEE, Robotics Industry Association) developing effective voluntary guidelines and standards.
- Government agencies (e.g., FDA, FAA, NHTSA) regu- lating in ways that foster innovation.
- Non-governmental organizations certifying companies and products (e.g., Underwriters Laboratories, Consumer Reports).
- Accounting firms (e.g., KPMG, EY, Deloitte, PwC) pro- viding auditing services to ensure trustworthiness.
- Insurance companies compensating for failures to guar- antee trustworthiness.

II. EMBRACING THE HUMAN-CENTERED APPROACH: ENHANCING CONTROL AND AUTOMATION

Adopting the Human-Centered Artificial Intelligence (HCAI) mindset emphasizes strategies to enable users to maintain control over highly automated systems. By focus- ing on interdisciplinary teamwork, intuitive user interfaces, and careful design considerations, HCAI empowers users to steer, operate, and control automated devices effectively. It emphasizes the importance of human users feeling in control, understanding the technology, and being able to intervene when necessary.

A. *Encouraging Interdisciplinary Collaboration*

Successful designs enable humans to work collaboratively in interdisciplinary teams, promoting coordination and collab- oration with various stakeholders.

B. *Designing Intuitive User Interfaces*

Well-designed user interfaces play a vital role in reducing workload, enhancing performance, and ensuring safety [7]. They provide necessary support for human activities, allowing users to intervene and control the technology effectively.

C. *Addressing Special Cases:*

In instances requiring either fully automatic action or com- plete human control, additional design considerations are nec- essary. These situations demand tailored approaches to strike the right balance between automation and human intervention [8].

D. *Broadening Horizons: Achieving Beyond RST Goals*

ISSN Online : 2709-5088
ISSN Print : 2709-507X
IJCST

The HCAI framework presented in Section 3 offers a roadmap for designers and researchers to explore new possibil- ities in advancing RST systems. Beyond RST, the framework supports broader goals such as ensuring privacy, enhancing cybersecurity, promoting social justice, and preserving the environment. By aligning technology with human aspirations and simplifying users' efforts, the HCAI framework paves the way for transformative advancements in various domains [9].

III. THE TWO-DIMENSIONAL HCAI FRAMEWORK

This section introduces the two-dimensional HCAI frame-
work, a departure from the traditional one-dimensional think- ing of human control versus computer automation [10]. It explores the four quadrants:

- Lower Left Quadrant (Low Computer Automation, Low Human Control): Simple devices with minimal automa- tion, such as clocks or music boxes.
- Upper Left Quadrant (High Human Control, Low Au- tomation): Human autonomy, where individuals derive pleasure from mastering activities like baking or playing musical instruments.
- • Lower Right Quadrant (High Computer Automation, Low Human Control): Computer autonomy requiring rapid action, like airbag deployment or pacemakers, with no time for human intervention.
- Upper Right Quadrant (High Computer Automation, High Human Control): Desired goal of Reliable, Safe & Trust- worthy (RST) systems, involving complex tasks with creative decisions and standardized contexts.

IV. PROMETHEUS PRINCIPLES: A FRAMEWORK FOR DESIGN EXCELLENCE

Introduce the Prometheus Principles, a set of guidelines that emphasize comprehensible, predictable, and controllable interfaces [11]. These principles include:

- Consistent Interfaces: Users can form, express, and revise intent through consistent interfaces.
- Continuous Visual Display: Users receive continuous visual feedback on objects and actions of interest.
- Rapid, Incremental, and Reversible Actions: Users can perform actions quickly, incrementally, and reverse them if needed.
- Error Prevention: Designs include safeguards to prevent errors.
- Informative Feedback: Users receive informative feed- back for every action, acknowledging their input.
- Progress Indicators: Systems show progress indicators to indicate the status of ongoing processes.
- Completion Reports: Users receive reports confirming the accomplishment of tasks.

A. Examples of Human-Centered Automation

- Automobiles, Explore the evolution of automobile au- tomation, from power steering to adaptive cruise con- trol, highlighting the balance between human control and automation, emphasizing the need for active driver supervision.
- Home Appliances, Describe how home appliances like dishwashers and ovens provide users with control set- tings, visual feedback, and automation features, enhanc- ing user experience and ensuring efficient operation.
- Examine how cameras in smartphones offer users the ability to express their intent visually, with continuous feedback on captured images and options for further ad- justments, demonstrating a balance between automation and user control.
- Patient Controlled Analgesia (PCA) Devices, Explain how PCA devices offer different levels of

ISSN Online : 2709-5088
ISSN Print : 2709-507X
IJCST

automation and human control, from basic morphine drip designs to advanced RST designs incorporating sensors [12], machine learning, and hospital control centers.

SUMMARY

The Human-Centered Automation and Intelligence (HCAI) framework, discussed in the paper, distinguishes the roles of human control and computer automation. It emphasizes the importance of well-designed automation, enabling human operators to maintain control while benefiting from advanced technological capabilities. The framework outlines scenarios where rapid computer-driven actions are essential, highlights the significance of human mastery, and cautions against exces- sive automation or human control. It suggests leveraging the unique strengths of both humans and computers for optimal system performance.

LIMITATIONS

While the HCAI framework provides valuable insights, it also faces limitations. One major challenge is the absence of concrete, objective measures for control and autonomy across diverse tasks, hindering precise design discussions. Ad- ditionally, ethical concerns such as responsibility, fairness, and explainability are acknowledged but not fully addressed, leav- ing gaps in the practical implementation of these principles. The framework underscores the complexities of achieving a balanced approach to automation and human control, including addressing potential de-skilling effects and vigilance issues as user actions decrease.

CONCLUSION

In conclusion, the HCAI framework offers a multidimen- sional perspective on automation and human control, guiding the design of technologies for enhanced user performance. It advocates for the recognition of both human and computer capabilities, fostering innovation and improvement. However, the framework's effectiveness relies on the development of objective control measures and the resolution of ethical chal- lenges. Despite these limitations, the framework represents a significant step toward creating reliable, safe, and trustworthy systems that empower users while incorporating advanced automation. Further research and practical implementation are necessary to fully realize the potential of the HCAI framework in shaping the future of artificial intelligence and technology.

REFERENCES

[1] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Col- lisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R.,& Horvitz, E. (2019). Guidelines for human-AI interaction. Proceedings of the 2019 CHI Conference on Human.

[2] Venigandla, K., & Tatikonda, V. M. (2021). Improving Diagnostic Imaging Analysis with RPA and Deep Learning Technologies. Power System Technology, 45(4).

[3] Factors in Computing Systems (Paper 3, pp. 1-13). ACM.

[4] Apple Computer, Inc. (2019). Human Interface Guidelines. Cuper- tino, CA: Apple. https://developer.apple.com/design/human-interface- 1guidelines/ios/overview/themes/.

[5] Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), 775–779. Pergamon. https://doi.org/10.1016/0005-1098(83)90046-8.

[6] Bennett, K. B., & Hoffman, R. R. (2015). Principles for interaction design, Part 3: Spanning the creativity gap. IEEE Intelligent Systems, 30(6), 82–91. https://doi.org/10.1109/MIS.2015.108.

[7] Berry, J. C., Davis, J. T., Bartman, T., Hafer, C. C., Lieb, L. M., Khan, N., & Brilli, R. J. (2016). Improved safety culture and teamwork climate are associated with decreases in patient harm and hospital mortality across a hospital system. Journal of Patient Safety, 1.

[8] Blackhurst, J. L., Gresham, J. S., & Stone, M. O. (2011). The autonomy paradox. Armed Forces Journal, 20–40. http://armedforcesjournal.com/the-autonomy-paradox/.

[9] Bradshaw, J. M., Hoffman, R. R., Woods, D. D., & Johnson, M. (2013).The seven deadly myths of autonomous systems.

[10] Canadian Government. (2019). Responsible use of artificial intelligence (AI). Canada.ca. https://www.canada.ca/en/government/system/digital- government/modern-emerging-technologies/responsible-use-ai.html.

[11] Candy, L. (2020). Creating with the digital: Tool, medium, me- diator, partner. In A. L. Brooks & C. Sylla (Eds),

ISSN Online : 2709-5088
ISSN Print : 2709-507X
IJCST

Interactivity, game creation, design, learning, and innovation. Springer (to appear). http://lindacandy.com/wp-content/uploads/2019/12/ArtsIT-LCandy.pdf.

[12] Defense Science Board. (2016). Summer study on autonomy. Office of the Undersecretary for Defense for Acquisition, Technology and Logistics, Department of Defense. Du, M., Liu, N., & Hu, X. (2020). Techniques for interpretable machine learning. Communications of the ACM, 63(1), 68–77. https://doi.org/10.1145/3359786.

[13] Dudley, J. J., & Kristensson, P. O. (2018). A review of user interface de- sign for interactive machine learning. ACM Transactions on Interactive Intelligent Systems (Tiis), 8(2), 8. https://doi.org/10.1145/3185517.