|  |
| --- |
| **Page 1** |

**Date received** :

**Corresponding author** : Xu Wei, xuwei11@zju.edu.cn

Human-AI Interaction: A new interdisciplinary field that realizes the concept of "Human-centric AI"

Xu 1 , GE Liezhong 1 , high peak 2

(1. Psychological Science Research Center, Zhejiang University, Hangzhou 310058 ; 2. Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou 310058)

Abstract: AI technology has benefited mankind and also brought challenges to research and development. If it is not developed properly, it will harm mankind and society. At present, there is no systematic cross-border

Disciplinary work framework to effectively respond to these new challenges. In order to comply with the crossover trend of disciplinary development, the National Natural Science Foundation of China established a crossover in 2020.

Ministry of Science. In this context, this article analyzes the new challenges faced by the research and development of AI systems, and further elaborates on the "human-centered AI" we proposed in 2019.

(HCAI) R&D philosophy and design goals. The HCAI research and development concept is currently one of the hot topics in the AI ​​industry abroad. In order to promote the implementation of the HCAI concept, we

The new interdisciplinary field of Human-Artificial Intelligence Interaction (HAII) is systematically proposed, and its purpose, scope, research and application focus are defined. Through literature review

And analysis, this article summarizes the focus of the research and application of HAII at home and abroad, and puts forward the main research directions in the future. Finally, for the future HCAI concept and HAII

In the field of work, a series of countermeasures and suggestions have been put forward.

Keywords: artificial intelligence; human-artificial intelligence interaction; autonomy; human-centered artificial intelligence; human-computer interaction; human factors engineering; human-AI system interaction;

Human-centered design

Chinese Library Classification Number: TP3-05

Document mark code: A

Article ID:

**Human-AI interaction:**

**An emerging interdisciplinary domain for enabling human-centered AI**

XU Wei 1 , GE Liezhong 1 , GAO Zaifeng 2

(1. Zhejiang University, Center for Psychological Sciences, Hangzhou 310058, China;

2. Zhejiang University, Department of Psychology, Hangzhou 310058, China)

**Abstract** The new characteristics of AI technology have brought new challenges to the research and development of AI systems. AI

technology has benefited humans, but if improperly developed, it will harm humans. At present, there is no systematic interdisciplinary

approach to effectively deal with these new challenges. This paper analyzes the new challenges faced by AI systems and further

elaborates the “Human-Centered AI” (HCAI) approach we proposed in 2019. In order to enable the implementation of the HCAI

approach, we systematically propose an emerging interdisciplinary domain of "Human-AI Interaction" (HAII), and define the objective,

methodology, and scope. Based on literature review and analyses, this paper summarizes the main areas of the HAII research and

application as well as puts forward the future research agenda for HAII. Finally, the paper provides strategic recommendations for

future implementation of the HCAI approach and HAII work.

**Keywords** artificial intelligence, human-artificial intelligence interaction, autonomy, human-centered artificial intelligence, human-

computer interaction, human factors engineering, human-AI system interaction, human-centered design

|  |
| --- |
| **Page 2** |

1 Introduction

Artificial intelligence (AI) technology is benefiting mankind, but the goal

The research and development of many previous AI systems mainly follow the "technology-centric"

Idea [1-5] . Research shows that inappropriate AI technology development has led to

Many accidents that harm humans, the AI ​​accident database has collected 1,000

There have been many accidents [6] , these accidents include pedestrians killed by self-driving cars,

Errors in trading algorithms led to a "flash crash" in the market, and facial recognition systems led to

Innocent people were arrested and so on. Academician of the American Academy of Engineering, Professor of Computer

Shneiderman [4] will focus on "technology-centric" or "people-oriented

The controversy in developing an AI system for the “centered” concept is visually described as

"AI Copernicus Revolution", proposed that AI development should put humans in the middle

Heart, not algorithms and AI technology.

In recent years, how to focus on the concept of “Human-centric AI”

Prevent AI from harming humans and causing negative social impacts, etc.

Research has attracted more and more attention [2-3][5-11] , currently there is no

There is a systematic interdisciplinary work framework to effectively deal with these

New challenges and promotion of work in this area. National Natural Science of China

The Science Foundation Committee established an interdisciplinary department in 2020,

At the high-end academic forum of Fork Science, invited AI and human-computer interaction

(Human-computer interaction), human factors engineering (human

factors engineering) and other experts agree that interdisciplinary

It is the inevitable trend of future scientific development.

In this context based on the concept of interdisciplinary cooperation, this article

Answer the following three questions: Compared with traditional computing technology, AI technology

What new challenges does technology bring? How to promote "people-centered

The application of “AI” concept in AI research and development? From the perspective of interdisciplinary cooperation

What strategy should we adopt? This article will further elaborate on our

The “human-centered AI” (human-centered AI) proposed in 2019

AI, HCAI) concept [2] , systematically proposed human-artificial intelligence interaction

(Human-AI interaction, HAII) this emerging interdisciplinary

field. It is hoped that through advocating the HCAI concept and the HAII field,

AI research and development benefit mankind and avoid potential negative effects.

2 New changes and challenges brought by AI technology

2.1 Cross-age characteristics of AI technology

The AI ​​community generally believes that AI technology has mainly experienced three waves.

The first two waves focused on scientific exploration, confined to “technology-centered

The vision of the "heart" presents the characteristics of "academic dominance". The depth machine

Technologies such as computer learning, computing power, and big data have promoted the third wave of

Rise. In the third wave, people began to pay attention to the application of AI technology.

Use landing scenarios to develop front-end applications and human-computer interaction that are useful to humans

Mutual technology, considering issues such as AI ethics. At the same time, the AI ​​community

It is advocated to regard human and AI as a human-machine system, and introduce the role of humans [1][3] .

It can be seen that the third wave began to revolve around the "human factor".

The development of AI has prompted people to think more about the theory of “human-centered AI”

read. Therefore, the third wave presents "technology improvement + application development"

The characteristics of “development + people-centered” [2] means that AI development is not only

Is it a technical solution or a system engineering of interdisciplinary cooperation.

2.2 The new human-machine relationship in the intelligent era

AI can be developed into a feature of autonomy

Agent. Depending on the degree of autonomy, an AI system can have a certain

To a degree similar to human cognition, learning, self-adaptation, independent execution

Operation and other capabilities, in a specific scenario, it can autonomously complete a

Some specific tasks can be autonomous in some unanticipated scenarios

To complete the tasks that the previous automation technology could not complete [9-11] .

This kind of intelligent and autonomous feature gives the machine in the man-machine system new

character of. In the age of non-intelligence, human operations are based on computing technology,

The machine acts as an auxiliary tool. The nature of interaction between humans and AI systems

The above is the interaction between humans and autonomous agents. With the improvement of AI technology,

It is possible for the main agent to use an auxiliary tool that supports human operations

Develops into a teammate working with a human operator, pretending to be

Play the double role of "auxiliary tool + human-computer cooperation teammate" [11] .

Therefore, the human-computer relationship in the intelligent age is evolving into a team

The relationship between teammates forms a kind of "human-machine crew" (human-machine crew)

teaming) cooperation [13-14] . The human-machine relationship in the intelligent age

Different from the human-computer interaction in the PC era, it is a challenge and a machine for AI research and development.

In case, developers need to use this kind of human-machine cooperation in AI research and development,

Ensure that humans can effectively control the AI ​​system and avoid harming humans.

2.3 Comparison of human-non-AI system interaction and human-AI system interaction

Human-computer interaction is an interdisciplinary field formed in the PC era.

Investigate the interaction between people and non-AI computing systems. Table 1 compares people-

Some features between non-AI system interaction and human-AI system interaction

Levy. The characteristics of the human-AI system interaction are based on the AI ​​system

Has a high degree of intelligent autonomy, some features are not yet available

now. It can be seen from Table 1 that compared with human-non-intelligent system interaction, human

-Intelligent system interaction brings many new features and new problems, and also gives

The research and application of human-intelligent system interaction has brought new opportunities.

In the human-non-intelligent system interaction, as a way to support human

Operational auxiliary tools, machines rely on pre-designed rules and calculations

Law. Although there is also a certain degree of human-machine cooperation between humans and machines,

But the machine as an auxiliary tool is passive, and only humans can dominate

Start this limited cooperation actively .

The autonomous characteristics of the AI ​​system agent make the agent

It is similar to a human-human team to a certain extent.

"Cooperative interaction" between. In a specific operating environment, this

This kind of interaction can be two-way active, shared, mutual

|  |
| --- |
| **Page 3** |

3

Complementary, replaceable, adaptive, goal-driven and predictable

Determined by the characteristics of the test (see Table 1). With the development of AI technology

In the future, AI systems will have more of these characteristics [12] .

This shows that the new characteristics of human-AI system interaction in the intelligent age

And the research issues have far exceeded the current human-computer interaction research

The scope of research and application requires a new thinking to consider how to

Effectively carry out multidisciplinary cooperation to deal with human-AI system interaction and

A series of new features and challenges faced in the research and development of AI systems.

3 The rise of the HAII field and the field concept

Foreign research and applications for human-AI system interaction have been developed

Open [13-14] . For example: human-agent interaction [15] , human-autonomous interaction

(human-autonomy interaction) [16] , human-AI interaction [17] .

Although these tasks have their own focus, they are all researches on people and intelligence

The interaction between "machines" (agents, intelligent agents, etc.). so,

This kind of interaction is essentially human-AI interaction (human-AI

interaction, referred to as HAII). There is no system yet

Regarding the work framework in the field of HAII, it is necessary to formally promote the

Promote as a new interdisciplinary field.

3.1 The concept of the HAII field: human-centered AI

In recent years, when the “technology-centric” approach has influenced AI research

At the same time as the development, researchers are also exploring

AI development methods, for example, human-centered algorithms, AI humanistic design

Design, inclusive design, AI based on social responsibility [8] .

Stanford University established "Human-centric AI" in 2019

Table 1 Comparison of the characteristics of human-non-AI computing system interaction and human-AI system interaction

**Table 1 Comparative analysis between human interaction with non-AI systems and AI systems**

feature

Human-non-AI computing system interaction in the computer age

Human-AI system interaction in the intelligent age

Instance

Office software, washing machine, automatic production line, etc.

Intelligent audio, intelligent decision-making system, self-driving car, etc.

Machine behavior and intelligence are determined according to fixed algorithms, logic and rules.

Machine behavior; lack of machine intelligence

Have varying degrees of human-like cognitive abilities (learning, self-adaptation, self-holding

Line, etc.); show special and evolvable machine behavior

Machine role

Mainly as an auxiliary tool

May also become teammates working with humans

Machine output

Deterministic

Uncertain

Human operator role monitor, performer

May also become teammates working with AI (people should be the final decision makers)

Man-machine relationship

Human-computer interaction

Human-computer interaction + human-crew team cooperation

User Interface

Graphical user interface, touch screen interaction, explicit interaction

Wait

Also includes new intelligent interaction: voice interaction, face recognition

No, brain-computer interface, implicit interaction, etc.

Human-computer interaction

Behavioral characteristics

Human-computer interaction initiated by humans and based on explicit interfaces is based on information such as human cognition, behavior, emotion, and scene context. Agents can also

To actively initiate human-computer interaction based on the implicit interface, etc.

Starting ability

Humans initiate tasks and actions actively, and machines passively accept that both human and machine can actively initiate tasks and actions

Human-computer interaction

Directionality

Only people’s one-way trust and situational awareness of machines

Knowledge, decision-making, etc.

Two-way trust, situational awareness, intention between human and machine

Figure, the decision-making control rights that can be shared between humans and machines (humans should

Have ultimate control)

Smart complementarity

Machines have no intelligence, and there is no intelligence complementation between humans and machines. The complementarity between machine intelligence and human biological intelligence

System output

Interpretability

Mainly depends on the availability of the system output interface

Also presents AI "black box" effect, leading to system output

Difficult to explain and understand

Human, machine

Ability to predict

Only human operators have

Both humans and machines can use models such as behavior and situational awareness,

Predict the other party's behavior, environment, system, etc.

Adaptive ability

Only human operators have

Both humans and machines can adapt to each other's behavior and operating scenarios

Goal setting ability

Only human operators have

Both human and machine can set or adjust the system target

Replacement ability

Machines can replace human tasks (with the help of automation

Technology, mainly physical)

Machines can replace human physical and cognitive tasks (man-machine can take the initiative

Or passively take over, delegate tasks, etc.)

Human-machine cooperation

limited

Based on some of the above characteristics, it may produce more effective human-machine cooperation

User needs

Mainly include usability, psychology, safety, physiology, etc., but also emotion, privacy, ethics, decision-making autonomy, etc.

|  |
| --- |
| **Page 4** |

(Human-Centered AI, hereinafter referred to as HCAI) Research Center,

The purpose is to develop a cooperative

It is about human morals and ethics and AI systems that benefit humans [18] .

Xu Wei [2][19] proposed a "people-centered

AI” (HCAI) system conceptual framework, which includes human, ethical,

Three aspects of technology. Shneiderman [20] proposed a

A guiding frame for the development of a reliable, safe and trustworthy AI system

shelf. HCAI is the concept that guides the new field of HAII. Below us

Further elaborate the three aspects of the HCAI concept: technology, people, ethics

Management (see Figure 1). Among them, Figure 1 summarizes the main ways of all aspects of work.

Diameter (see the blue font surrounding the three peripheral circular parts in Figure 1),

For example, human needs, AI application scenarios; Figure 1 also summarizes these

HCAI design goals to be achieved by the work (see Figure 1

Human-centered AI" the black font in the center circle part), for example,

Available AI, useful AI.

(1) "Technology": Emphasize the organic combination of the three parts.

(A) Machine intelligence: use algorithms, big data, computing power and other technologies to

Develop machine intelligence; (b) Human intelligence: use intelligence enhancement technology,

Promote the increase of human intelligence with the help of methods such as psychology and cranial nerve technology

Strong (see 4.1); (c) Human-machine hybrid enhanced intelligence: AI community has recognized

Realizing that the path of independent development of AI technology has encountered a bottleneck effect,

High-level human cognition is difficult to reach the level of human intelligence

[1][3][20] . Therefore, the HCAI concept emphasizes the integration of human roles into humans and machines

System, through the complementation of man-machine intelligence, develop man-machine hybrid enhanced intelligence

The integration of energy, AI and human intelligence enhancement technology (see 4.1, 4.2).

The goal is to develop sustainable, powerful, and human-controllable AI;

The purpose of AI development is to enhance human abilities, not to replace humans.

(2) "Human" aspect: Emphasizes

Starting from needs, implement effective application scenarios, and develop human cognitive models

Type, implement human-computer interaction based on “human-centered” in AI R&D

Mutual design and methods (modeling, design, testing, etc.). The purpose is to develop

Useful (meet people’s needs, have use value), usable (easy

Use, easy to learn ), humans have the final control of the AI ​​system.

(3) "Ethics": Combining interdisciplinary methods and effective development

Development practices, standards and governance, etc., through engineering design methods (such as

"Meaningful human control", see 4.7), to ensure that AI development complies with

Based on fairness, human privacy, ethics, human decision-making power, etc.

Rights. The goal is to develop ethical and responsible AI.

The HCAI concept emphasizes maintaining the central position of people in AI development,

A systematic AI development that implements the interdependence of technology, people, and ethics

Develop thinking and advocate that AI development is an interdisciplinary collaborative system engineering

Process, developed a reliable, safe and trustworthy AI system [2][20] .

3.2 Framework of work in the HAII field

For the emerging field of HAII, we have made the following preliminary

Definition. Figure 2 illustrates the leadership of Human-Artificial Intelligence Interaction (HAII)

Domain framework, in which the blue circle represents the main cross-disciplinary collaboration

As a subject, the long white part represents the research of HAII discussed in this article.

The main problems of research and application (see 4.1 to 4.7 for details).

• HAII field concept: Human-centered AI (HCAI).

• HAII field purpose: as an interdisciplinary field,

HAII uses AI, computer science, human-computer interaction, human factors engineering,

Psychology and other disciplines technology and methods, committed to cooperative research and development of the AI ​​department

System, optimizing the interaction between humans and AI systems, focusing on machines and humans

Complementary advantages of intelligence, comprehensive consideration of AI ethics, emphasizing people

|  |
| --- |
| **Page 5** |

5

The ultimate control over the AI ​​system, by providing an interdisciplinary

Cooperation platform to realize HCAI development concept in AI system development,

Provide humans with safe, reliable, and trustworthy AI.

• HAII research and application scope: in a narrow sense, HAII involves

To the research and application of human interaction with AI systems; broadly speaking, any

AI research and development involving human use and influencing people belong to HAII

Scope, including the research and application of AI systems that interact with people

Application areas, such as smart phone application APP, smart human-computer interaction technology,

Intelligent decision-making system, intelligent Internet of things, etc. As shown in Figure 2, HAII

From the perspective of the human-machine-environment system, considering various factors

The impact of AI interactions, a comprehensive understanding of these impacts will help play AI

Advantages of technology to avoid negative effects.

• HAII domain approach: as an interdisciplinary field, through

Multidisciplinary approach (modeling, algorithm, design, engineering, testing, etc.)

And the process of cooperation to develop the AI ​​system. These methods come from these

Related disciplines, such as computational models, engineering design methods, behavioral disciplines

Learn research methods, human-computer interaction design, etc.

• HAII field personnel: those who are engaged in the research and application of HAII

The staff includes people from AI, computer, data science, human-computer interaction, mind

Professionals in science, cognitive neuroscience, social science, etc. Broadly

Generally speaking, most people engaged in the research and development of AI systems belong to this

In this category, the AI ​​systems they develop more or less interact with people,

All are to develop AI systems that are beneficial to humans.

3.3 Why new areas of HAII are needed

First, the HAII field provides a cooperation platform for various disciplines.

tower. HAII helps to unite under one domain name (HAII)

Interdisciplinary and cross-industry professionals involved in the research and development of AI systems, avoid

Confusing names, this kind of interdisciplinary cooperation helps to effectively develop

Human-centric AI system.

Secondly, the field of HAII emphasizes that the objects of its research and application are

AI, not a traditional non-AI system, helps remind people to pay attention

The difference in characteristics between AI and non-AI systems has prompted people to pay attention to AI

Use effective methods to solve new challenges and new problems brought about by technology

Solve the unique problems in the development of AI systems.

Finally, the HAII field helps to promote the HCAI concept in AI research

Implementation in development. HAII emphasizes the central role of generals in AI research and development

Integrate into the system design to avoid potential security risks [8][12] .

Historically, new technologies have promoted the emergence of new fields. Enter PC

Times, the traditional "human-machine interaction" (human-machine interaction)

Interaction) field transition to the new version of "human-computer interaction" (human

-Computer interaction), but this "machine" is not the other "machine" [21] . intelligent

The machine of the era transitions to the AI ​​system, and the new characteristics of AI prompt HAII

The emergence of new fields, therefore, the emergence of HAII is also inevitable.

We do not recommend setting up HAII as an independent new school

Section, emphasize that HAII is a new type of interdisciplinary field, and hope that through this

Cross-disciplinary and cross-industry collaboration in various fields to implement the HCAI concept.

For example, the work of HAII requires the participation of human-computer interaction personnel, and they must

New thinking must be adopted to carry out research and application on the interaction between humans and AI.

3.4 Challenges in realizing the HCAI concept and HAII solutions

Work in the HAII field is not just emerging. The AI ​​industry and others

|  |
| --- |
| **Page 6** |

Some work has been carried out in related disciplines. To further elaborate on HAII

Field, based on literature review and analysis, Table 2 summarizes the current HAII

The focus of research and application, the challenge of realizing the HCAI concept, HAII

Possible solutions in the field and expected HCAI design goals. Table 2

Some examples are also cited, and the details are discussed in Part 4 of this article.

As can be seen from Table 2, first of all, in order to realize the design based on the HCAI concept

As planned, there are many challenges to be solved in the research and application of HAII. Do not

To solve these challenges, we will not be able to realize the HCAI concept, and we will not be able to develop

Send out safe, reliable, and trustworthy AI.

Secondly, Table 2 outlines the scope of research and application of HAII.

Professionals from AI and other disciplines are currently working on this

Work, these challenges cannot be solved by a single discipline.

It shows that the work of HCAI and HAII requires interdisciplinary collaboration.

Finally, the HAII field adopts an interdisciplinary approach. HAII research

And application challenges, possible solutions and examples all depend on

Interdisciplinary methods (modeling, engineering design, behavioral science methods, etc.),

A single-disciplinary approach cannot effectively solve these problems.

More than 40 years ago, when the new PC technology was just emerging, the developer base

This book follows the "technology-centric" philosophy. With the popularity of PC

And, many user experience problems have arisen, and people are beginning to realize

The importance of the concept of "human-centered design" comes from the computer science

Professionals in science, human factors engineering, and psychology have promoted the human-machine

The formation and development of interactive disciplines. Years of practice, user experience

The idea has formed a consensus in society and computing technology circles.

Today, with the introduction of AI technology, we have witnessed similar

Scenario, but this time ignoring the cost of the "people-centered" concept

The impact on humans and society will be more serious [6] . Therefore, the various disciplines

We must once again work together to promote the "people-centered

"Heart Design" version (i.e. HCAI concept) and work in the field of HAII,

Make more effective use of AI technology, maximize strengths and avoid weaknesses, and serve humanity.

In addition, the HCAI concept and the HAII field define AI development

The concepts, goals, and approaches that should be followed in

Body model or algorithm. We emphasize interdisciplinary cooperation, once it is clear

With these ideas, goals, and approaches, AI personnel can be more effective

Develop models, algorithms, and

technology.

4 Key issues in the research and application of HAII

According to the HCAI concept, in response to the new challenges posed by current AI,

We analyze the current progress in the research and application of HAII from the following aspects:

Development and put forward the key directions for the future. At present, some AI personnel are also opening

We hope that the HCAI concept and the HAII field will be proposed

Can strengthen the HCAI concept and interdisciplinary cooperation of these personnel,

It is also hoped that non-AI personnel will actively participate in the research and application of HAII.

4.1 Human intelligence enhancement

Since the concept of AI was proposed in 1956, researchers have started

Committed to another path: Intelligent enhancement (Intelligence

Augmentation, IA) [22] . Intelligence enhancement is committed to enhancing human intelligence

Can [23-24] . Researchers use new technologies (such as psychology, brain-computer interface

Mouth, virtual reality), and use AI technology to promote the research of intelligent enhancement

Research and application [52] . From the perspective of HCAI concept, the

Strong has the same goal: use AI technology to enhance human abilities.

There is a long-standing competition between the AI ​​field and the intelligent enhancement field.

Some AI personnel believe that AI can replace humans, and intelligence enhances humans

Members believe that AI only provides new means for intelligent enhancement technology [25] .

From the analysis of the HCAI concept, AI and intelligence enhancement will adopt similar technologies.

It should be the expansion of human intelligence, and it should be "using human

"Centered" partnership, many smart solutions are actually two technologies

Collection, haii can act as a bridge for cooperation between the two.

There are many problems in intelligent enhancement research that require the leadership of HAII in the future.

Domain contribution. First of all, machine intelligence cannot imitate certain aspects of human intelligence.

In these dimensions, HAII advocates exploring which type of

Human intelligence enhancement and technology can provide effective means to compensate

The weakness of AI [26] , which requires intelligent augmentation personnel to actively seek

AI, psychology, cognitive neuroscience, human factors engineering discipline support.

Secondly, develop the application of the best combination of AI and intelligent enhancement technology

The solution will effectively promote the cooperation between the two technologies, thereby

Achieve the effect of "1 + 1> 2". The HAII field can play a

The role of an intermediate bridge, from the perspective of human-computer interaction, psychology, etc.

Degree of human-computer interaction, multi-modal interaction compatibility, human recognition

Know the processing level, the degree of autonomy of the AI ​​system and other dimensions to develop

Solutions that can support the effective interaction between humans and intelligent systems.

Third, according to the HCAI concept, intelligent enhancement professionals should

Human beings are placed at the center of the system plan. The HAII field promotes psychology,

Cognitive neuroscience and other personnel actively participate in research. For example, based on

Plasticity mechanism, build a controllable cognitive load, timely physiological feedback, physical

A new type of human-computer interaction study of two-way brain interaction. These studies will be effective

Supports training and enhancement of human intelligence in many application fields [53] .

Finally, look for AI technology and intelligence at the biological neural level

Integrated solutions for enhanced technology. This is the current research side

One direction, for example, brain-computer fusion [27] . HAII promotes AI and cognitive God

Optimized through the cooperation of science, brain imaging technology, human-computer interaction and other personnel

Brain-computer interface solutions, through the integration of AI at the biological neural level

Explore effective human-AI interaction methods with intelligent enhancement technology [28] .

4.2 Human-machine hybrid enhanced intelligence

Introducing the role of humans and human intelligence into the AI ​​system will form

The advantages of human intelligence and machine intelligence complement each other to develop a stronger

Large and sustainable human-machine hybrid enhanced intelligence [1][3][20][28] .

|  |
| --- |
| **Page 7** |

7

Table 2 Challenges in realizing the HCAI concept in the HAII field, possible HAII solutions, application examples, and HCAI design goals

Figure 2 Challenges in realizing the HCAI design philosophy, possible HAII solutions, application examples, and expected HCAI design goals

HAII field focus

(See section 4)

Challenges in realizing the HCAI concept

Based on the HCAI concept

HAII possible solutions

Research and application examples in the field of HAII

Based on the HCAI concept

Design goals (Figure 1)

AI system

Machine behavior

(Section 4.7)

Potential, biased system output, unexpected

Machine behavior, unique machine behavior evolution, not

Mature machine learning training and testing methods, lacking

Machine learning without user participation, social interaction

Complex machine behavior, between multiple AI agents

Complex behaviors and interactions

"Human-centric" machine learning, interactive

Machine learning, human-computer interaction methods in data collection

Application in collection, training, algorithm adjustment, and testing

Use, machine behavior based on behavioral science methods

Research

Interactive machine learning: [29] Based on the "human-centered" concept, the target users

(Domain experts) Participate directly, reduce reliance on machine learning experts, and pass

Build and train machine learning models through effective human-computer interaction. User check,

Training model results, and continuously adjusting subsequent inputs until satisfactory results are obtained. Mutually

For traditional machine learning, this method is faster, more efficient and optimized, and it has been applied

Used in recommender systems, information retrieval, situational awareness and other fields

Useful AI,

Ethical AI,

Responsible AI

Human intelligence enhancement

(See section 4.1)

Human Intelligence Enhancement Technology (IA) and AI Technology

The competition between IA and AI technology is lacking

The best solution with complementary advantages

HAII can play a role as a bridge between IA and AI

Use, formulate the best combination of the two technologies

Plan to ensure that AI technology enhances human intelligence,

The guarantor has the ultimate control right to develop biological

Neural-level solutions (brain-computer fusion, etc.)

An intelligent system that integrates IA and AI technologies to enhance human intelligence in applications,

Through monitoring (remote, etc.), humans can realize human-machine collaboration and guarantee human’s

Decisive control [25] . For example, smart drones (military, etc.), robots (monthly

Ball detection and other dangerous scenes, medical and surgical fine operations, etc.), intelligent decision-making system

(Stock trading system, etc.)

Useful AI,

Available AI,

Enhance people's abilities,

People have final control

Man-machine hybrid

Strong intelligence

(See section 4.2)

Machine intelligence is difficult to simulate human advanced cognitive abilities

Power, the bottleneck effect of the development of machine intelligence technology,

Lack of development paths to develop machine intelligence in isolation

Sustainability, lack of human control in AI development

And safety considerations

Human-machine hybrid enhances intelligence, "Human in the loop"

AI system and interactive design, man-machine collaborative control

System, brain-computer hybrid system, based on cognitive psychology

Cognitive computing (emotion, intention

Etc.), human advanced cognitive ability model, knowledge

Consciousness representation and map, symbiosis and fusion of man and machine

"Human in the loop" hybrid enhanced intelligence [1] [30-31] : the use of human and machine intelligence

Complementary advantages, dealing with large-scale, incomplete and unstructured knowledge and information;

In the interaction between users and the AI ​​system, iterative knowledge and learning will deepen the understanding of data and system.

The system understands that the AI ​​model accepts specific input and determines the input based on user feedback.

To achieve better results than those achieved separately, and avoid the loss caused by AI technology.

Risk control, which has been applied in fields such as autonomous driving, auxiliary medical care, and video retrieval

Scalable AI,

Efficient and powerful AI,

Human-controllable AI,

People have final control

People-AI cooperation

(See section 4.3)

Lack of mature people-AI cooperation theory, method,

Cognitive framework, lack of mature human-machine situational awareness

Sharing, mutual trust between humans and machines, and shared mental models between humans and machines

Theories, models and methods of sharing, human-machine decision-making and sharing

Law

Human-AI cooperation theory and model, human-AI cooperation

For team performance evaluation and testing methods, man-machine

Co-driving, human-AI cooperation in the social environment,

Human-computer interaction model of human-AI cooperation, human-made

Designed for the ultimate decision-maker's person-AI cooperation

From a multi-disciplinary perspective, it is based on people from the perspective of perception, cognition, and execution-

AI solutions provided by AI cooperation provide the basis. For example, human-AI collaborative research

Strategy and framework [32] , human-AI cooperation team performance evaluation [33] , human-AI mutual trust

[34] , Mental model in human-AI cooperation [35] , System design of human-AI cooperation [36] ,

Human-AI cooperation system permissions [37] , quantitative and qualitative modeling of human-AI cooperation [38]

Useful AI,

Available AI,

Human controllable AI

Explainable AI

(See section 4.4)

AI "black box" effect, users cannot understand AI

System decision-making, influence human decision-making, influence AI

"Human-centered" interpretable AI, terminal

User-participatory interpretable AI, understandable

"Human-centered" real-time interpretable AI solution: the computer in [39]

In game research, the real-time generation of AI agent and user interaction is based on natural language

Available AI,

Responsible AI

|  |
| --- |
| **Page 8** |

Technology promotion, psychological explanation theory has not been applied

Use, there is no way for end users to participate, AI

It is interpretable but not understood by the end user

AI, AI system output interface visual design

Design, "people in the loop" interpretable AI, heart

The transformation and application of Neo-Confucianism interpretation theory, user participation

And interactive AI man-machine interface design

Inference data, use these data to train an AI model that can

The result of the play generates human intelligible reasoning. In the study of [40], autopilot

Passengers in the driving car select driving environment targets through effective human-computer interaction methods,

These selection goals make the decision of the driving algorithm easier to interpret and understand

Human controllable autonomy

(See section 4.5)

AI's unique autonomy characteristics and potential negatives

Impact on the mixing of the concepts of automation and autonomy

Confusion, the underestimated impact of autonomy (confused as

High level of automation)

Mankind is controllable and autonomous, and man and machine share autonomy

, Human-autonomous teamwork, automation

Transformation of research results, meaningful human control

System, autonomous fault tracking data system

Udelv Automatic Delivery Vehicle (ADV) [41] , ADV adopts L4 automatic driving, which can

Complete unmanned point-to-point operations in some public places. The system

Including remote monitoring system, if needed (mission marginal area, emergency scene

Etc.), through seamless human-machine control conversion, manual intervention can be achieved

Available AI,

Ethical AI,

Human-controllable AI,

People have final control

Intelligent human-computer interaction

(See section 4.6)

Lack of human-computer interaction paradigm for intelligent interaction,

Human limited cognitive resources in a complex intelligent computing environment

The bottleneck effect of the source, for the non-AI system

Human-computer interaction design standards, the use of intelligent interaction

User experience problem

New paradigm of intelligent human-computer interaction, effective intelligence

Human-computer interaction design, aimed at people with AI systems

Computer interaction design standards, based on human-AI cooperation

Human-computer interface design, intelligent human-computer interaction

Usability design

Human-computer interaction design standards for AI systems: ISO human-AI system interaction

Technical documents (ISO 9241-810) [42] , Microsoft Design Guidelines and Guidelines [17] ,

"Google AI + People Guidebook" [43] ; effective smart man-machine

Interactive development process: "paired AI cooperative development process" [44] ; effective people

Computer interaction design method: AI priority method [43] , AI as design material [45]

Available AI,

Useful AI

Ethical AI design

(See section 4.7)

Humans may lack ultimate control over AI systems

Right, the AI ​​system produces output deviations and unexpected results

As a result, abuse of the AI ​​system (resulting in discrimination, privacy

Leakage, etc.), lack of traceability of AI system failures

And accountability mechanisms

Meaningful human control, AI error traceability machine

System, transparent design, optimized machine learning construction

Model, training, testing, AI personnel knowledge improvement

Ascend, the AI ​​ethical design of an interdisciplinary approach

Design, ethical design techniques and examples

"Meaningful human control" design [46] : transparent system design, effective human

Machine interaction, human operators can have enough information about the autonomous technology used

(Situational awareness, etc.) to ensure informed, conscious and legal decisions are made;

Equipped with a “fault tracking system” mechanism to realize system behavior and fault tracing and questioning

Accountability

Ethical AI,

Responsible AI,

Available AI,

Human controllable AI

|  |
| --- |
| **Page 9** |

9

At present, the research on hybrid augmented intelligence can basically be divided into two

kind. The first type is the "people in the loop" hybrid at the system level

Enhanced intelligence [57] . This kind of thinking is in line with the HCAI concept, which is about human creation

Introduce into the AI ​​system to form a human-centered and human-machine relationship

Hybrid intelligence. For example, in the "people in the loop" paradigm, people always

It is a part of the AI ​​system. When the confidence of the system output is low, people dominate

The parameters are adjusted automatically to give a reasonable and correct problem solution, which constitutes an improvement

Intelligent feedback loop [1] [48] . Another option is at the biological level

The development of "brain in the loop" hybrid augmented intelligence [3] , based on biological intelligence

With the goal of deep integration with machine intelligence, through neural connection channels,

In order to form an enhancement, substitution and compensation for a certain functional body.

The second type of hybrid augmented intelligence is to embed human cognitive models into AI

In the system, a hybrid enhanced intelligence based on cognitive computing is formed [1] . From

HCAI concept analysis, this kind of hybrid enhanced intelligence is not really meaningful

The human-machine hybrid on the Internet enhances intelligence, because this kind of system is not able to guarantee

Use the human-machine system as the carrier to realize the central role of humans in the human-machine system

Use and final decision-making power. Of course, introduce the human cognitive model to the machine

In intelligence, it is very important for the development of machine intelligence.

Work in the HAII field will contribute to the research and application of human-machine hybrid enhanced intelligence.

Play an important role. First, the HAII field promotes psychology, cognition

Cooperative support from professionals in engineering and other disciplines, accelerating existing psychology, etc.

The conversion of academic achievements to support cognitive computing research and provide effective

Cognitive computing architecture [49] , for example, to improve the understanding of unstructured audiovisual

The ability to perceive information and the processing efficiency of massive amounts of heterogeneous information,

The HAII field needs to support the AI ​​community’s “perception feature extraction and expression

And integration" and "modal information collaborative computing" and other aspects of audiovisual

Research on Information Cognitive Computing [50] . HAII encourages the use of cognitive computing

Law and the "people in the loop" approach (systems, biological level) integration

Work Ideas. Based on the HCAI concept, this kind of thinking helps to develop

A more powerful, sustainable, human-controllable AI.

Second, carry out the human-machine hybrid intelligent control based on the HCAI concept

Research. For the control of human-machine hybrid intelligent system, there are currently two main types

Scheme: "Human in the loop control" and "Man-machine coordinated control" [51] . AI

Efficient switching between humans and machines in the emergency state of the system is currently important

Research topics. For example, the tracking control of an autonomous weapon system after launch,

Efficient man-machine switching under emergency conditions for autonomous driving vehicles. HCAI philosophy

Humans are required to have the ultimate control, which requires specialized expertise such as AI and human-computer interaction.

The cooperation of industry personnel to find effective solutions.

Third, carry out research on human-computer interaction of human-computer hybrid enhanced intelligent systems

Research. "Human-in-the-loop" hybrid intelligent systems need to interact with users

Mutual design [30] . Different from traditional human-computer interaction, the object of user interaction is

The AI ​​model, the user interface is difficult to understand, and the user interacts with the AI

Uncertainty of user intent [1] . HAII research requires AI professionals

Cooperate with human-computer interaction, human factors engineering, etc., from intelligent system, application

Optimize system design in three aspects: household and human-computer interaction design. For example, open

Develop natural interaction design and choose an effective mental model.

Finally, carry out the human-machine hybrid increase at the human advanced cognitive level

Strong intelligence research. HAII research requires AI, cognitive neuroscience, computing

Cooperation of professionals in computer science, psychology, etc. For example, enter one

Further explore the research of human-computer fusion, brain-computer fusion, etc., in the future,

At a higher cognitive level, it is the superposition of brain-computer intelligence (such as learning,

Memory) to establish more effective models and algorithms [52] ; to explore how to integrate people

Decision-making and experience and machine intelligence in logical reasoning, deductive reasoning, etc.

The combination of advantages in this aspect makes human-machine cooperation highly efficient [1] . Long-term

From a long-term perspective, human-machine hybrid enhanced intelligence may form an effective human-machine in the future

Symbiosis [53] , through the intelligent integration of individuals and groups, and other ways, finally in

Realize the symbiosis and integration of man and machine at the system and biological level [54] .

4.3 Human-AI cooperation

Intelligent technology has brought a new type of human-machine relationship: human-AI cooperation

The work of the human-AI system as a combination is better than the work of a single entity

More effective [33][35] . Research on human-AI cooperation is currently

Hotspots. HAII research requires AI, psychology, and human factors engineering

The cooperation of science and technology is carried out on the three levels of perception, cognition, and execution.

At the perceptual level, for effective human-AI cooperation, the AI ​​system

Human models are needed to support the system’s monitoring of human status (health

Theory, behavior, emotion, wish map, ability, etc.); the man-machine of the AI ​​system

The interface should be transparent enough to help humans understand the current system state. example

For example, situational awareness (situational awareness) sharing between humans and machines is a combination of human-AI

One of the basic questions for research. Research needs to understand how to effectively implement

Two-way communication between the present human and AI based on the situational awareness model [55] ,

There is still a lack of situational awareness models and tests for human-AI cooperation

Method [66] . The future work of HAII needs to enrich the theory of situational awareness,

Provide support for human-AI cooperation modeling, cognitive architecture, and performance measurement.

At the cognitive level, the human-computer interaction model in the PC era has not

It can meet the complex interactive scenarios of the intelligent age. HAII requires a construct

Cognitive and computational model of human-AI cooperation [57] . Between people and AI

Mutual trust affects the performance of human-AI cooperation, HAII needs to study trust measurement

Measurement, modeling, repair, correction, etc., and how to measure

When the dynamic function exchange between human and machine in different operation scenarios

The trust needed. Different from traditional human-computer interaction, both human and AI need each other

This perceives and recognizes the intent and emotion of the partner, and future research needs to be further

Explore mental models, intention recognition, emotional interaction and other models step by step to

And how to implement and verify these models in system design.

At the executive level, effective human-AI cooperation should allow for

At the level of services, functions, systems, etc., the decision-making power is between humans and AI agents.

Sharing between management. The transfer of decision-making power depends on the two-way communication between man and machine.

Responsibilities, situational awareness sharing, degree of cooperation and other factors. For example, in

In the field of driving vehicles, the work of HAII needs to study the sharing of human-machine control

|  |
| --- |
| **Page 10** |

Sharing of situational awareness, mutual trust between humans and machines, and risk

Evaluation, etc., to ensure the rapid and effective switching of vehicle control rights between man and machine

Change, to ensure that people have the ultimate control (including remote control, etc.) [58] .

HAII research needs to understand how to complete between humans and machines under what conditions

Effective handover, can we use the idea of ​​human-AI cooperation?

Effective human-computer interaction, providing effective human-machine control transfer. HAII

Future work should also be carried out in the following aspects.

First of all, HAII needs to develop new theories for human-AI cooperation research,

Models and methods for evaluating and predicting the performance of human-AI cooperation teams, which

These are new problems that are not encountered in traditional human-computer interaction. HAII collar

Domain work should support AI modeling and the demand for modeling data (examples

For example, situational awareness, behavior, intention, trust) [13] , co-developers

-AI cooperation solutions in various application fields.

Secondly, the field of HAII needs to be

Research on human-AI cooperation on a level. To study social factors (social responsibility

Responsibilities, ethics, etc.) on the impact of human-AI cooperation on how to make AI replace

Rationally take on the role of a team and cooperate with human teammates [13] ; how to study

From the perspective of system design, develop well through appropriate human-computer interaction

Good human-computer relationship (trust, emotion, etc.); research how to effectively support

Support long-term human-AI cooperation (such as medical robots).

Third, the people needed for HAII research to build a human-AI cooperation scenario

Machine interaction modeling. Collaborate with AI personnel to research on humans-AI in cooperation with humans

Computer interaction modeling poses challenging theoretical issues, for example, distributed cognition

Theory, context-based knowledge representation and knowledge graph [59] , human-AI

Cognitive modeling of emotional interaction and social interaction in cooperation.

Finally, HAII needs to research in a real operation and social environment.

Research people-AI cooperation [60] . For example, laboratory studies have shown that compared with simple machines

The interaction of robots can enhance human coordination, and the robot can directly

Cooperating with people [61] ; the cognitive style of AI and people in human-AI cooperation,

When personality characteristics and other characteristics are adapted, it can enhance the mutual trust and reliability of humans and machines.

In the future, we will verify these human-AI social interactions in a real social environment,

This will help optimize the design of human-AI cooperation.

4.4 Explainable AI

Methods such as deep learning will produce AI "black box" effects, leading to

Users have questions about the decision-making of the AI ​​system. This effect can be used in various applications.

Occurs during use, including AI in financial stocks, medical diagnosis, and security inspections.

Monitoring, legal judgments and other fields, resulting in a decrease in the efficiency of system decision-making and ethical

Ethics and other issues affect the public's trust in AI.

Seeking explainable AI (eXplainable AI, XAI) has become

A research hotspot in the AI ​​community, the representative one is DARPA [62] in

Projects launched in 2016. The project focuses on: (1) Development or improvement

ML technology to obtain interpretable algorithm models; (2) With the help of advanced

Human-computer interaction technology to develop user interface models that can explain AI; (3)

Evaluate psychological interpretation theories to help explain the development of AI.

After years of hard work, the AI ​​community has begun to realize that non-AI disciplines are

It can explain the importance of AI research [63] . Miller [64] 's questionnaire

It is clear that most interpretable AI projects are only carried out within the AI ​​discipline. May

Many AI personnel adopt an “algorithm-centric” approach, which intensifies the calculation

The opacity of the law [65] . Some AI personnel did not follow the HCAI philosophy,

It is common to build interpretable AI for oneself rather than for users [66] . Research also

Show that if you adopt behavioral science methods, focus on users rather than

Technology, research on interpretable AI will be more effective [64] .

The future work in the field of HAII mainly includes the following aspects. head

First, research and develop "human-centered interpretable AI" solutions

case. The HAII field should start from human-computer interaction, psychology, human factors engineering, etc.

To find a solution. Many previous studies were based on static and

One-way information transmission-style interpretation, in the future, the work of HAII needs to study and explore

So-type, natural, interactive interpretation to design an interpretable interface [67-68] .

Second, the HAII field advocates that interpretable AI research needs to be further

Mining models such as psychology. Although these theories and models are usually

Produced in laboratory research, explainable AI research should be used well

These models also verify their feasibility [68] . HAII's work

Can use its own interdisciplinary characteristics to serve as an intermediate bridge

Function, accelerate the conversion of the theory, and construct an effective interface or calculation model.

Finally, in the field of HAII, the research and application of understandable AI should be carried out.

use. Explainable AI should also be understandable [68] , interpretability is a must

A condition is necessary, but not a sufficient condition. Considering from the HCAI concept, it is reasonable

The AI ​​solution should meet the needs of the end user (e.g. knowledge level).

Research in this area requires the support of behavioral science methods and experiments

Verify [69] .

4.5 Human controllable autonomy

Intelligent autonomous technology is entering people’s work and life,

However, some people have begun to confuse the concepts of automation and autonomy.

Can lead to inappropriate expectations and misuse of technology [10] . Automation technology

According to fixed algorithms, logic and rules, a certain machine behavior is generated.

Intelligent systems will have varying degrees of human-like intelligence (self-adaptation

Ability to respond, self-execution, etc.), the output of the system is uncertain, with

Machine behavior that may be deviating. The autonomous characteristics of AI are important for safety

Negative effects such as public psychology have not attracted enough attention [70] .

Starting from the HCAI concept, we advocate

The human-controllable and autonomous design of the “AI” design goal, that is, the intelligent self

The main system needs human monitoring, and the human operator has the ultimate

The decision-making power (through direct or remote control, etc.) [4] .

The human factors engineering community has already

Etc.) has carried out extensive research on the automation system in

Know [71-72] . Many complex automation systems have vulnerabilities.

It runs well in the specified operating scenarios, but unexpected things happen

It may cause the operator's "automation surprise" phenomenon [71] :

The operator cannot understand what the automation is doing and why it is doing it.

Endsley [70] believes that in the intelligent autonomous system, as the system

|  |
| --- |
| **Page 11** |

11

The improvement of the degree of system autonomy, the "automation" level of each individual function

Increase accordingly, which may cause operators to pay attention to these functions

In the emergency state, the phenomenon of "people outside the ring" appears. right

Investigations of fatal accidents of autonomous vehicles that have occurred in recent years show that

Confusion of face patterns, “people outside the ring”, and excessive trust are precisely based on the

Typical problems in automation systems [73-74] .

Features such as potential autonomous operability in smart technology will also cause

People have an excessive trust in this technology similar to automation. have

The autonomous system of learning ability means the uncertainty of its operation results

Sex may develop in unexpected ways, and it may bring the operator

A more intense "automation surprise" experience than automation.

Future work in the field of HAII can be considered from the following aspects:

First of all, the HAII field must address some basic issues of autonomy

Start research. Fully understand the characteristics of AI autonomy from the perspective of human-computer interaction

Impact on the design of human-computer interaction, study the expectations of autonomy on operators,

Roles, etc., to study the influence of autonomy on the emotional stress and cognition of operators

The influence of ability, personality traits and communication attributes [75-76] .

Secondly, in the field of HAII, HCAI should be realized in the development of independent systems

The conceptual design goal of "Human Controllable AI". Currently, despite the human-computer interaction

Mutual professionals participate in autonomous systems (such as self-driving cars)

Research and development, but frequent accidents remind us to evaluate the current

Law [11] . SAE [77] believes that L4-L5 autonomous vehicles do not require people

Category monitoring and intervention, we questioned SAE’s neglect of automation and autonomy

The essential difference between the [78] , may be the design, safety, standardization and recognition

Certificate, etc. have an adverse effect. High-level self-driving cars are

Dynamic "autonomous system, not a traditional automation system. Based on

The concept of HCAI is based on human-AI cooperation, mutual trust between humans and machines, and situational awareness.

Exploring independent design from the perspectives of sharing, independent sharing, etc., to achieve effective

Human-machine co-driving and handover [79] , any level of autonomous vehicles needs to confirm

The guarantor is the ultimate controller of the system (including remote control methods).

Finally, to achieve "meaningful human control" for autonomous systems

"Meaningful human control" (meaningful human control) design goal [46] .

The HAII work must implement this goal: (1) Through the "Human in the Environment

"Road" and human-computer interaction design to ensure that humans can take over or

Interrupt the operation of the system; (2) Install the

Tracking System" to achieve design improvement and accountability system for man-machine faults [46] ,

Promote the realization of the design goal of "Human Controllable AI" in the HCAI concept.

4.6 Intelligent human-computer interaction

Intelligent human-computer interaction has brought challenges and opportunities to the HAII field.

Meet. The rich application scenarios and user needs of the AI ​​system need to be effective

Human-computer interaction paradigm [80] . Limitations of existing human-computer interaction methods (such as WIMP)

Due to limited sensing channels, insufficient interaction bandwidth, input/output bandwidth

Issues such as imbalance and unnatural interaction methods have been put forward by research

Post-WIMP paradigm [81] , these paradigms require the verification of HAII’s work

certificate. Multi-modal fusion and parallel interaction paradigm are important for HAII in the future

Research content, research in this area is currently mainly in the AI ​​and computing technology circles

Expanding, HAII should provide interdisciplinary support. HAI also needs to be in context

Greater breakthroughs have been made in perception and intention understanding [81-83] .

The new human-computer relationship of human-AI cooperation improves human-computer interface design.

A new request was made. Traditional man-machine interface is mainly based on "stimulus-response"

The concept of "instruction sequence" interaction is aimed at intelligent human-computer interaction (emotional

Sense, intent recognition, context detection, etc.)

Mutual, HAII must design an effective human-machine interface to support human-AI cooperation

Required situational awareness sharing, human-machine mutual trust, human-machine control sharing, etc.

The HAII field needs to develop human-computer interaction design for AI systems

standard. The existing standards are mainly for non-AI systems, international standards

The organization is drafting the design standard for the interaction of the AI ​​system (ISO

9241-810) [42] , there have been some human-computer interaction devices for AI systems

Planning guidelines [17][64] , but the contribution of the HAII field is still needed.

4.7 Ethical AI Design

The ethical AI design goal advocated by the HCAI concept is the AI ​​industry

Important issues of general concern at present. Research shows that AI personnel are on the job

There is usually a lack of training in the application of ethical codes for design in industry training.

The AI ​​community has realized that ethical AI design requires multidisciplinary cooperation

[84] . Professional organizations and companies have issued multiple sets of AI ethics codes, but

Research shows how to effectively implement these regulations in the development of AI systems

Fan needs further efforts. Some professionals are

Not to consider ethical design in the process [85] . Therefore, in the field of HAII

An important task is to implement ethical AI design in the development process

middle.

First of all, the HAII field should carry out research on the behavior of AI machines.

Research. In 2019, many scholars from MIT and other universities published in Nature

The article suggests to carry out research on AI machine behavior [9] . Currently engaged in machinery

For the research, it is mainly AI personnel who have not received any training in behavioral sciences.

The non-determinism of AI machine behavior needs to be based on algorithms, data, training,

Research the influencing factors in aspects such as testing to avoid algorithm deviation. Currently,

There have been researches on machine behavior based on the HCAI concept, for example, "Human

Centered machine learning", interactive machine learning and other methods [29][66] .

These methods help to solve the extreme behavior of AI system,

Fairness and other issues.

Second, the HAII field can adopt interdisciplinary methodology to support

Ethical AI design, iterative design advocated by human-computer interaction

And the evaluation method is used in the model algorithm training, and the collection algorithm training

Data, which defines the expected results of users and converts them into valid inputs

According to data, using early prototypes to carry out user experience evaluation, through iterative

Design and test to reduce algorithm deviation [86] .

Third, adopt the method of "meaningful human control" to transform ethics

Intelligent AI is implemented in the system design [46][70] . The system design must ensure:

(1) The operator can be informed about the autonomous technology used

And make a conscious decision; (2) the operator has enough information to ensure

|  |
| --- |
| **Page 12** |

Take legal action. In addition, in system design, testing, professional

Personnel training and other aspects to ensure the effective control of the system by humans.

system. "Meaningful human control" is consistent with the HCAI concept and is beneficial to

Realize the goal of ethical AI in design.

Finally, the HAII field needs to optimize AI development practices to support the

Physical and chemical AI design. Research shows that the current AI community lacks effective ethics

Improved AI design methodology, technical details or details to guide design options

Example [85] . The field of HAII can start from "people in the loop" AI and human-AI cooperation

Provide help with design ideas. How to embed ethical AI principles

Into the development process, how to improve the ethical AI standards for AI engineering

The impact of teacher behavior, how to improve the ethical design skills of AI engineers

The problems of ability and other aspects all require HAII's multidisciplinary program.

5 Challenges and countermeasures of HAII research and application

As an emerging field, in the practice of realizing the HCAI concept,

HAII research and application must face challenges.

First, AI personnel lack an understanding of the HCAI concept. Many AI

The personnel in the development basically follow the "technology-centric" concept, one

Some people think that the problems that human-computer interaction cannot solve are currently being used by AI technology.

Technical solutions (such as voice input), the man-machine interface does not have to consider user experience;

Professionals such as human-computer interaction often define product requirements in AI projects.

Participating in the project later, limiting their influence on the design of the AI ​​system,

And it led to the failure of some AI projects [87-89] .

Second, there is a lack of effective interdisciplinary cooperation in the research and development of AI systems.

Lack of effective communication language between AI personnel and non-AI personnel, non-AI personnel

People lack the necessary AI knowledge, and AI people lack knowledge of other

Section’s understanding [90] . At present, researches have put forward the concept based on HCAI

Of AI and human-computer interaction professionals’ "paired AI collaborative development

Process” [46] and improved human-computer interaction methods [91] .

Third, there is a lack of effective methods for interdisciplinary cooperation. Some man-machine interaction

In the development of AI systems, mutual personnel still use traditional non-AI systems.

Traditional methods; many AI personnel are not easy to accept methods from other disciplines.

At present, people have proposed some methods, for example, "AI first"

Method, "AI as a design material" [43][45] .

In order to effectively realize the HCAI concept and carry out the work of HAII, we start from

Recommendations are made at the following three levels (as shown in Figure 3).

First, at the AI ​​R&D team level, establish a multidisciplinary team and

Adopt an interdisciplinary approach. The only new problem that AI brings is the use of multi-discipline

Only through cooperation can we find effective solutions. Based on the HCAI concept to optimize the existing

AI R&D process, for example, formulating HCAI design goals in the early stages of development

Optimize collaboration and cooperation on various development process nodes.

Secondly, at the organizational level of AI R&D companies, cultivate HCAI-based

Conceptual organizational culture, formulate development standard guidelines based on HCAI,

Provide HCAI R&D resources (interdisciplinary human resources, HAII projects,

Interdisciplinary research equipment, etc.), establish an efficient AI R&D team.

Finally, at the social level of AI R&D, cultivate HCAI

Interdisciplinary compound talents who want to study. For example, in colleges

Undergraduate majors of "Major + Minor" and "Major + AI Minor", training acupuncture

Master and Ph.D students on key issues of HAII; formulate AI development strategies and methods

Regulations, etc.; carry out cross-industry and cross-disciplinary research projects, and promote academic circles

Collaboration with industry; set up a special government fund to support

HAII project; establish a complete scientific research system that combines production, education and research,

Carry out research and application of HAII in some key industries.

6 Conclusion

1. The current intelligent age presents "technology improvement + application

Development + people-centered" stage characteristics. Have a certain degree

The ability of learning, self-adaptation, independent execution, etc. is the unique

Special autonomy features give the machine a new role in the human-machine system

Color brings a new type of human-computer relationship: human-AI cooperation. With people-computing

Compared with computer interaction, human-AI interaction has brought many new features and problems.

Question, which brings new challenges to the research and development of AI systems, prompting us to re-evaluate

Estimate the current R&D strategy based on non-AI systems.

2. In order to effectively solve the new challenges in the research and development of AI systems, this article

Explains the concept of HCAI, which emphasizes the organic integration of the three aspects of work

combine. (1) "Technology": The development strategy that promotes AI technology should focus on

The advantages of human and machine intelligence complement each other, bringing the role of humans into the AI ​​system

System, resulting in a more powerful and sustainable human-machine hybrid increase

Strong intelligence. The purpose of AI is to improve people's abilities, not to replace people;

(2) "Human": Emphasizes starting from the needs of human beings, providing useful,

Available, human beings have the final decision-making power of AI system; (3) "Ethics":

From the perspective of human ethics, AI system research and development should avoid ethical ethics

Germany and other issues, the purpose is to develop ethical and responsible AI.

3. In order to realize the implementation of HCAI concept in AI research and development, this

The text system proposes the new interdisciplinary field of HAII. HAII field is available

In order to provide an effective cooperation platform for various disciplines, focusing on AI technology

The new challenges and new problems of technology help to promote the HCAI concept in AI

The implementation of research and development has developed a safe, reliable and trustworthy AI system

System.

4. Aiming at the new problems brought by AI, we propose the future of HAII

The focus of research and application, they include human intelligence enhancement technology

Technology, human-machine hybrid enhanced intelligence, human-AI cooperation, interpretable AI,

Human controllable autonomy, intelligent human-computer interaction, and ethical AI design.

|  |
| --- |
| **Page 13** |

13

5. In response to the challenges that HAII will face in practice in the future, from the R&D team

On the three levels of team, R&D organization, and R&D social environment, we recommend:

Establish a multidisciplinary R&D team, adopt an interdisciplinary approach, and optimize existing AI

R&D process; cultivate the organizational culture of HCAI concept, develop HCAI standard

Quasi-guidelines and R&D resources; cultivating interdisciplinary AI talents, political

The government has issued relevant policies and regulations, and actively carried out cross-disciplinary and cross-industry

And the research and application of HAII in key industries. We look forward as long as

All disciplines and industries work together, a "people-centered

The era of intelligent society based on the concept of AI will come.

references

[1] ZHENG N, LIU Z, REN P, et al. Hybrid-augmented

intelligence: collaboration and cognition[J]. Frontiers of

Information Technology & Electronic Engineering, 2017,

18(2):153-179.

[2] XU W. Toward human-centered AI: A perspective from

human-computer interaction[J]. Interactions, 2019,

26(4): 42-46.

[3] WU Z. Hybrid Augmented Intelligence. 2017, 2019.

[http://www.kejilie.com/iyiou/article/BjMBRb.html](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=http://www.kejilie.com/iyiou/article/BjMBRb.html)

[https://www.wicongress.org/2020/zh/article/487](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=https://www.wicongress.org/2020/zh/article/487) (video)

(Wu Zhaohui, Hybrid Augmented Intelligence, 2017, 2019)

[4] SHNEIDERMAN B. Design lessons from AI's two

grand goals: Human emulation and useful

applications[J]. IEEE Transactions on Technology and

Society 2020, 1(2): 73-82.

[5] SHNEIDERMAN B. Bridging the gap between ethics

and practice: Guidelines for reliable, safe, and

trustworthy Human-Centered AI Systems[J]. ACM

Transactions on Interactive Intelligent Systems

(TiiS), 2020, 10(4): 1-31.

[6] MCGREGOR S. AI Incidents Database

[https://incidentdatabase.ai/ ,](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=https://incidentdatabase.ai/) accessed on July 8 th 2021.

[7] XU W. User-centered design (V): From automation to the

autonomy and autonomous vehicles in the intelligence

era[J]. Chinese Journal of Applied Psychology, 2020,

26(2): 108-129.

[8] AUERNHAMMMER, J. (2020) Human-centered AI: The

role of human-centered design research in the development

of AI [M], in Boess, S., Cheung, M. and Cain, R. (eds.),

Synergy – DRS International Conference 2020, 11-14

August, https://doi.org/10.21606/drs.2020.282

[9] RAHWAN I, CEBRIAN M, OBRADIVUCH N, et al.

Machine behavior[J]. Nature, 2019, 568(7753): 477-486.

[10] KABER D B. A conceptual framework of autonomous and

automated agents[J]. Theoretical Issues in Ergonomics

Science, 2018, 19(4): 406-430.

[11] XU W. From automation to autonomy and autonomous

vehicles: Challenges and opportunities for human-

computer interaction[J]. Interactions, 2021, 28(1):49-53.

[12] Xu Wei, Ge Liezhong. Engineering Psychology in the Intelligent Age[J]. Psychological Science

Progress, 2020, 28(9): 1409-1425.

XU W, GE L. Engineering psychology in the era of

artificial intelligence[J]. Advances in Psychological

Science, 2020, 28(9): 1409-1425.

[13] BRILL JC, CUMMINGS ML, EVANS III AW,

HANCOCK PA, LYONS JB, ODEN K. Navigating the

Advent of Human-Machine Teaming[C]. In Proceedings

|  |
| --- |
| **Page 14** |

of the Human Factors and Ergonomics Society Annual

Meeting, Los Angeles, USA, 2018, 62(1): 455-459.

[14] JOHNSON M, VERA A. No AI Is an Island: The Case for

Teaming Intelligence[J]. AI Magazine, 2019, 40(1): 16-28.

[15] PRADA R, PAIVA A. Human-agent interaction:

Challenges for bringing humans and agents together[C]. In

Proc. of the 3rd Int. Workshop on Human-Agent

Interaction Design and Models (HAIDM 2014) at the

13th Int. Conf. on Agent and Multi-Agent Systems

(AAMAS 2014), 2014: 1-10.

[16] CUMMINGS ML, CLARE A S. Holistic modelling for

human autonomous system interaction[J]. Theoretical

Issues in Ergonomics Science, 2015, 16(3): 214-231.

[17] AMERSHI S, WELD D, VORVOREANU M, FOURNEY

A, …HORVITZ E. Guidelines for human-AI interaction.

CHI 2019, May 4-9, Glasgow, Scotland, UK, 2019.

[18] LI F F. How to make AI that's good for people. The New

York Times, 2018, March 7 th .

[19] XU W. User-centered Design (IV): Human-centered

artificial Intelligence[J]. Chinese Journal of Applied

Psychology, 2019, 25(4): 291-305.

[20] SHNEIDERMAN B. Human-Centered Artificial

Intelligence: Reliable, Safe & Trustworthy[J].

International Journal of Human–Computer Interaction,

2020, 36:(6): 495-504.

[21] Xu Wei. Discussion on new ideas for the research and application of human-computer interaction.

Ergonomics, 2005 , 11(4): 37-40.

XU W. Recent trend of research and applications on

human-computer interaction[J]. Chinese Journal of

Ergonomics, 2005, 11(4): 37-40.

[22] ENGELBART D C. Augmenting Human Intellect: a

conceptual framework [R]. SRI Summary Report AFOSR-

3223, Air Force Office of Scientific Research, Washington

DC, 1962.

[23] RAISAMO R, RAKKOLAINEN I, MAJARANTA P,

SALMINEN K, J, FAROOQ A. Human augmentation:

Past, present and future[J]. International Journal of

Human-Computer Studies, 2019, 131-143.

[24] STANNEY KM, SCHMORROW DD, JOHNSTON M,

FUCHS S, JONES D, HALE KS, AHMAD A, YOUNG

P. Augmented cognition: An overview. Reviews of

Human Factors and Ergonomics, 2009, 5(1), 195–224.

[25] HASSANI H, SILVA ES, UNGER S, TAJMAZINANI M,

FEELY M S. AI or Intelligence Augmentation (IA): What

Is the Future? [J] AI, 2020, 1: 143-155.

[26] Wang Dangxiao, Zheng Yilei, Li Teng, etc. Multi-mode for human intelligence enhancement

Mode Human-Computer Interaction. Science in China: Information Science, 2018, 48: 449-465.

WANG DX, ZHENG YL, LI T, et al. Multi-modal

human-machine interaction for human intelligence

augmentation[J]. Sci Sin Inform, 2018, 48: 449-465.

[27] Wu Zhaohui, Yu Yipeng, Pan Gang, Wang Yueming. A review of brain-computer fusion systems

[J]. Life Science, 2014, 6, 645-649.

WU Z, YU Y, PAN G, WANG Y. A review of brain-

computer fusion systems[J]. Chinese Bulletin of Life

Sciences, 2014, 6, 645-649.

[28] STANNEY KM, SCHMORROW DD, JOHNSTON M,

FUCHS S, JONES D, HALE KS, AHMAD A, YOUNG

P. Augmented cognition: An overview. Reviews of

Human Factors and Ergonomics, 2009, 5(1): 195–224.

[29] AMERSHI S, CAKMARK M, KNOX WB, KULESZA T.

Power to the people: The role of humans in interactive

machine learning [J]. AI Magazine, Winter, 2014, 105-120.

[30] Hu Yuanda, Sun Xiaohua, Zhang Hexin, Zhang Shengchen, Yi Siqi. Supporters

Research on Interactive Design of Hybrid Intelligence in the Loop[J]. Packaging Engineering Art

Edition, 2020, 41(18), 38-47.

HU Y, SUN X, ZHANG H, ZHANG S. Interion Design in

Human-in-the-Loop Hybrid Intelligence[J] Packaging

Engineering, 2020, 41(18), 38-47.

[31] Zheng Qinghua. Preface to the topic of typical applications of human-machine hybrid enhanced intelligence[J]

Computer Research and Development, 2020, 57(12), 2479.

ZHENG Q. Preface to the applications of human-machine

hybrid enhanced intelligence[j]. Journal of Computer

Research and Development[J], 2020, 57(12), 2479.

[32] O'NEILL T, MCNEESE N, BARRON A, SCHELBLE B.

Human–autonomy teaming: A review and analysis of the

empirical literature [J]. Human

Factors, 0018720820960865.

[33] BANSAL G., NUSHI B., KAMAR E, LASECKI WS,

WELD DS, HORVITZ E. Beyond accuracy: The role of

mental models in human-AI team performance [C].

In Proceedings of the AAAI Conference on Human

Computation and Crowdsourcing, 2019, 7(1): 2-11.

[34] BINDEWALD JM, RUSNOCK CF, MILLER M E. (2017).

Measuring human trust behavior in human-machine teams

[C]. In International Conference on Applied Human Factors

and Ergonomics, 2017. Springer, Cham.

[35] BANSAL G., NUSHI B., KAMAR E, LASECKI WS,

WELD DS, HORVITZ E. Beyond accuracy: The role of

mental models in human-AI team performance [C].

In Proceedings of the AAAI Conference on Human

|  |
| --- |
| **Page 15** |

15

Computation and Crowdsourcing, 2019, 7(1): 2-11.

[36] OZKAYA I. The Behavioral Science of Software

Engineering and Human–Machine Teaming [J]. IEEE

Software, 2020, 37: 3-6.

[37] HARING KS, MOSLEY A, PRUZNICK S, FLEMING J,

SATTERFIELD K, ... FUNKE G. Robot authority in human-

machine teams: effects of human-like appearance on

compliance [C]. In International Conference on Human-

Computer Interaction 2019, 63-78. Springer, Cham.

[38] PETERS N. Interruption timing prediction via prosodic task

boundary model for human-machine Teaming[C]. In Future

Information and Communication Conference, 2019, 501-

522, Springer, Cham.

[39] EHSAN U, TAMBWEKAR P, CHAN L, HARRISON B,

RIEDL M O. Automated rationale generation: a technique

for explainable AI and its effects on human perceptions [C].

In Proceedings of the 24th International Conference on

Intelligent User Interfaces, 2019, 263-274.

[40] WINTERSBERGER P, NICKLAS H, MARTLBAUER T,

HAMMER S, RIENER A. Explainable Automation:

Personalized and Adaptive UIs to Foster Trust and

Understanding of Driving Automation Systems [C]. In 12th

International Conference on Automotive User Interfaces

and Interactive Vehicular Applications, 2020, 252-261.

[41] INTEL NEWS. Mobileye and Udelv Ink Deal for

Autonomous Delivery. 2021,

[https://www.intel.com/content/www/us/en/newsroom/news/](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=https://www.intel.com/content/www/us/en/newsroom/news/mobileye-udelv-deal-autonomous-delivery.html%23gs.15ib8a#gs.15ib8a)

[mobileye-udelv-deal-autonomous-delivery.html#gs.15ib8a](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=https://www.intel.com/content/www/us/en/newsroom/news/mobileye-udelv-deal-autonomous-delivery.html%23gs.15ib8a#gs.15ib8a)

[42] ISO (International Organization for Standardization).

Ergonomics-Ergonomics of Human-System Interaction –

Part 810: Robotic, Intelligent and Autonomous Systems

(version for review), 2020.

[43] GOOGLE PAIR. People + AI Guidebook: Designing

human-centered AI products. pair.withgoogle.com/ 2019

[44] CEREJO J. Design process of human-centered AI — Part2;

Bootcamp, 2021, https [://bootcamp.uxdesign.cc/human-](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=https://bootcamp.uxdesign.cc/human-centered-ai-design-process-part-2-empathize-hypothesis-6065db967716)

[centered-ai-design-process-part-2-empathize-hypothesis-](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=https://bootcamp.uxdesign.cc/human-centered-ai-design-process-part-2-empathize-hypothesis-6065db967716)

[6065db967716](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=https://bootcamp.uxdesign.cc/human-centered-ai-design-process-part-2-empathize-hypothesis-6065db967716)

[45] HOLMQUIST L E. Intelligence on tap: artificial

intelligence as a new design material [J]. Interactions,

2017, 24, 28–33.

[46] DE SIO S. DEN HOVEN VAN. Meaningful human

control over autonomous systems: A philosophical

account [J]. Front. Robot. AI 2018, 5: 15.

[47] ZANZOTTO F M. Human-in-the-loop Artificial

Intelligence[J]. Journal of Artificial Intelligence

Research, 2019, 64: 243-252.

[48] ​​PAN Y. Heading toward artificial intelligence [J]. 2.0.

Engineering, 2016, 2(4), 409-413.

[49] SALVI C, BRICOLO E, KOUNIOS J, et al. Insight

solutions are correct more often than analytic solutions[J].

Think. Reason., 2016, 22(4): 443-460.

[50] Zheng Nanning. The new era of artificial intelligence [J]. Intelligence Science and Technology

News, 2019, 1(1), 1-3.

ZHENG N. The new era of AI[J]. Chinese Journal of

Intelligent Science and Technology. 2019, 1(1), 1-3.

[51] Zhao Yunbo. Intelligent control of man-machine hybrid [M], Chapter 17. Wang Feiyue,

Chen Junlong (Editor-in-Chief), Intelligent Control. China Science and Technology Press, 2019.

ZHAO Y. Intelligent Control of Human-Machine

Hybrid[M], Chapter 17. In Wang F, Chen J (Editor),

Intelligent Control, Beijing: China Science and

Technology Press, 2019.

[52] WU Z, ZHENG N, ZHANG S, ZHENG X, GAO L, SU

L. Maze learning by a hybrid brain-computer [J].

system. Scientific reports, 2016, 6: 1-12.

[53] LICKLIDER R. Man-Computer Symbiosis[J]. Trans on

Human Factors in Electronics, 1960, (1): 4-11.

[54] SUN R, Potential of Full Human–Machine Symbiosis

Through Truly Intelligent Cognitive Systems[J]. AI &

Society, 2017: 1–12.

[55] KITCHIN J, BABER C. A comparison of shared and

distributed situation awareness in teams through the use of

agent-based modelling[J]. Theoretical Issues in

Ergonomics Science, 2016, 17(1), 8-41.

[56] STANTON NA, SALMON PM, SALAS E, HANCOCK

P A. State-of-science: situation awareness in individuals,

teams and systems [J]. Ergonomics, 2017, 60(4), 449-466.

[57] Liu Ye, Wang Yamin, Xuan Yuming. Human-computer cooperation psychology towards the intelligent age

Model[J]. Science in China: Information Science, 2018, 48(4), 376–389.

LIU Y, WANG YM, BIAN YL, et al. A psychological

model of human-computer cooperation for the era of

AI[J]. Sci Sin Inform, 2018, 48: 376–389.

[58] FRIDMAN L. Human-Centered Autonomous Vehicle

Systems: Principles of Effective Shared Autonomy. MIT

HCAV Research Program:

[https://arxiv.org/pdf/1810.01835.pdf,](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=https://arxiv.org/pdf/1810.01835.pdf) 2018.

[59] FAN JJ, TIAN F, DU Y et al. Thoughts on human-

computer interaction in the age of artificial intelligence (in

Chinese) [J]. Sci Sin Inform, 2018, 48, 361–375, doi:

10.1360/N112017-00221

|  |
| --- |
| **Page 16** |

[60] KRAFFT PM, MACY M, PENTLAND A. Bots As Virtual

Confederates: Design and Ethics[C]. in Proceedings of the

2017 ACM Conference on Computer Supported

Cooperative Work and Social Computing, 2017: 183–190.

[61] CRANDALL JW et al. Cooperating with machines[J]. Nat.

Commun. 2018, 9, 233

[62] GUNNING D. Explainable Artificial Intelligence (XAI) at

DARPA.

[https://www.darpa.mil/attachments/XAIProgramUpdate.pdf](https://translate.google.com/translate?hl=en&prev=_t&sl=zh-CN&tl=en&u=https://www.darpa.mil/attachments/XAIProgramUpdate.pdf)

. 2017.

[63] BOND R, MULVENNA MD, WAN H, FINLAY DD,

WONG A, KOENE A., ... & ADEL T. Human centered

artificial intelligence: weaving UX into algorithmic

decision making. In *RoCHI* 2019, 2-9.

[64] MILLER T, HOWE P, SONENBERG L. Explainable AI:

Beware of Inmates Running the Asylum, 2017. Available

online: https://arxiv.org/pdf/1712.00547.pdf

[65] EHSAN U, WINTERSBERGER P, LIAO QV, MARA M,

WACHTER S, ... RIEDL M O. Operationalizing Human-

Centered Perspectives in Explainable AI[C]. In Extended

Abstracts of the 2021 CHI, 2021, 1-6.

[66] KALUARACHCHI T, REIS A, NANAYAKKARA S. A

Review of Recent Deep Learning Approaches in Human-

Centered Machine Learning[J]. Sensors, 2021, 21(7): 2514.

[67] ADBUL A, VERMEULEN J, WANG D, LIM BY,

KANKANHALLI M. Trends and trajectories for

explainable, accountable and intelligible systems: an

HCI research agenda[C]. Proceedings of the 2018 CHI

Conference on Human Factors in Computing Systems

Montreal, QC, Canada, 2018, Paper No.582.

[68] BRONIATOWSKI D A. Psychological Foundations of

Explainability and Interpretability in AI, 2021.

[69] HOFFMAN RR, KLEIN G, MUELLER S T. Explaining

explanation for “explainable AI”[C]. In Proceedings of the

Human Factors and Ergonomics Society Annual Meeting,

2018, 62(1), 197-201. LA, CA: SAGE Publications.

[70] ENDSLEY M R. From here to autonomy: lessons learned

from human-automation research[J]. Human Factors,

2017, 59(1): 5-27.

[71] SARTER NB, WOODS D D. How in the world did we

ever get into that mode: Mode error and awareness in

supervisory control[J]. Human Factors, 1995, 37(1): 5-19.

[72] Xu Wei. Automated human-automation system interaction in the cockpit of an aircraft

Psychological Research[J]. Psychological Science, 2003, 26(3): 523-524)

XU W. A psychological study on the human-automation

interaction in civil flight deck[J]. Psychological Science,

2003, 26(3): 523-524.

[73] BAINBRIDGE L. Ironies of automation[J]. Automatica,

1983, 19(6): 775-779.

[74] XU W. Status and challenges: Human factors in

developing modern civil flight decks [J]. Chinese Journal

of Ergonomics, 2004, 10(4): 53-56.

[75] DE VISSER EJ, PAK R, SHAW T H. From automation to

autonomy: the importance of trust repair in human–

machine interaction[J], Ergonomics, 2018, 61: 1409-1427.

[76] CUMMINGS ML, BRITTON D. Regulating safety-critical

autonomous systems: past, present, and future perspectives

[M]. In Living with Robots, 2020, 119-140. Academic

Press.

[77] SAE (Society of Automotive Engineers). (2018).

Taxonomy and definitions for terms related to driving

automation systems for on-road motor vehicles [M].

Recommended Practice J3016 2018, revised 2018-06.

[78] HANCOCK P A. Some pitfalls in the promises of

automated and autonomous vehicles [J]. Ergonomics,

2019, 62, 479-495.

[79] JEONG K A. Human-system cooperation in automated

driving[J]. International Journal of Human-Computer

Interaction, 2019, 35(11): 917-918.

[80] XU W. User-centered design (II): New challenges and

opportunities[J]. Chinese Journal of Ergonomics, 2017, 23

(1): 82-86.

[81] Fan Junjun, Tian Feng, Du Yi, Liu Zhengjie, Dai Guozhong. Human Machines in the Intelligent Era

Some thoughts on interaction[J]. Science in China: Information Science, 2018, 48(4):

361–375.

FAN JJ, TIAN F, DU Y, et al. Thoughts on human-

computer interaction in the age of artificial intelligence[J].

Sci Sin Inform, 2018, 48: 361–375.

[82] Fan Xiangmin, Fan Junjun, Tian Feng, et al. Human-computer interaction and artificial intelligence: from

Alternate ups and downs to synergistic progress[J]. Science in China: Information Science, 2019,

49: 361-368.

FAN XM, FAN JJ, TIAN F, et al. Human-computer

interaction and artificial intelligence: from competition to

integration[J]. Sci Sin Inform, 2019, 49: 361-368.

[83] Ge Liezhong, Xu Wei. User Experience: Theory and Practice [M]. Beijing: Medium

Renmin University Press, 2020.

GE LZ, XU W. User Experience: Theory and Practice[M].

Beijing: China Renmin University Press, 2020.

[84] DONAHOE E. Human centered AI: Building trust,

democracy and human rights by design. An Overview of

|  |
| --- |
| **Page 17** |

17

Stanford's Global Digital Policy Incubator and the

XPRIZE Foundation's June 11th Event[C]. Stanford

Global Digital Policy Incubator (GDPi), 2018.

[85] MITTELSTADT B. AI ethics–too principled to fail?

2019, arXiv preprint arXiv:1906.06668.

[86] BUOLAMWINI J, GEBRU T, GENDER SHADES.

Intersectional Accuracy Disparities in Commercial Gender

Classification[J]. in Proceedings of the 1st Conference on

Fairness, Accountability and Transparency (eds. Friedler,

SA & Wilson, C.), 2018, 81: 77–91.

[87] YANG Q. Machine Learning as a UX Design Material:

How Can We Imagine Beyond Automation,

Recommenders, and Reminders[C]? Conference: 2018

AAAI Spring Symposium Series: User Experience of

Artificial Intelligence, March, Palo Alto, CA, 2018.

[88] YANG Q. STEINFELD A. ROSE C. ZIMMERMAN J.

Re-examining Whether, Why, and How Human-AI

Interaction Is Uniquely Difficult to Design[C]. In CHI

Conference on Human Factors in Computing Systems

(CHI '20), April 25–30, 2020, Honolulu, HI, USA. ACM,

New York, NY, USA, 2020: 12 pages.

[89] BUDIU R, LAUBHEIMER P. Intelligent Assistants Have

Poor Usability: A User Study of Alexa, Google Assistant,

and Siri[R]. 2018, https://www.nngroup.com

[90] GIRARDIN F, LATHIA N. When user experience

designers partner with data scientists[R]. In the AAAI

Spring Symposium Series Technical Report: Designing the

User Experience of Machine Learning Systems. 2017, The

AAAI Press, Palo Alto, California.

[91] XU W, GE L. New trends in human factors [J].

Advances in Psychological Science, 2018, 26(9):

1521-1534.

Xu Wei, PhD in Psychology and Computer Science in the US

Master's degree, professor; senior researcher of internationally renowned IT companies

Researcher, Chairman of Human-Computer Interaction Technical Committee. appoint

International Organization for Standardization (ISO) Human-System Interaction

Professional committee member, International

Journal of Human-Computer

Interaction and other SCI journal editorial board, China recognized

Director of the Engineering Psychology Branch and the Engineering Psychology Branch. Has been in China for more than 30 years

Foreign universities, well-known IT and civil aviation companies are engaged in human-computer interaction and human factors engineering

Aspects of research, design and standard development; presided over or participated in many countries, provinces and ministries

Level projects, won numerous research and design awards, and the results have been applied

Machine models and IT products; published 5 books in both Chinese and English,

Published in core academic journals such as interaction, psychology, aviation, computing technology, human factors engineering, etc.

Form a large number of papers; host or participate in the development of human factors engineering, human-computer interaction at home and abroad

More than 20 design standards.

Ge Liezhong, PhD, professor, worker of Chinese Psychological Society

The current chairman of the Cheng Psychology Professional Committee.

The main research directions are human-computer interaction, face recognition,

User usability and user experience research. Once presided over

8 national, provincial and ministerial projects; presided over Huawei and other companies

Division more than 30 horizontal projects; presided over international cooperation projects

Item 4 Published papers in academic journals at home and abroad

174 articles. Won the "Disciplinary Construction Achievement Award" of the Chinese Psychological Society in 2019, 2019

The leading team was awarded by the Central Military Commission’s Equipment Development Department and other 5 ministries

Collective Award for Outstanding Achievement in Space Manned Project".

Gao Zaifeng, PhD, Professor, Zhejiang University

Deputy Director of the Department of Physics and Behavioral Sciences, Changjiang Scholar

Young scholar, Chinese Society of Cognitive Ergonomics

Affair, editorial board member of "Applied Psychology" magazine. main

Engaged in cognitive psychology and engineering psychology

Relevant research and presided over national and provincial projects 9

Item; as the first author or corresponding author in

Psychological Science, Cognition, International

Journal of Human-Computer Interaction and other publications

More than 40 SCI/SSCI papers, serving as Automotive Innovation,

Cognition, Emotion, Cortex and more than 40 international core academic terms

Reviewer of the journal.