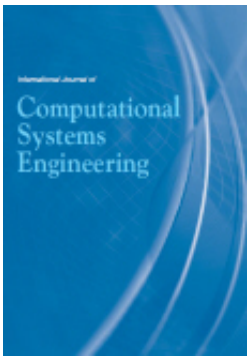
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Computational Systems Engineering

ISSN online: 2046-3405 - ISSN print: 2046-3391
https://www.inderscience.com/ijcsyse

Reproduction of humanness based on eXtended intelligence: concept of artificial personality and its mechanism

Takayuki Fujimoto

DOI: 10.1504/IJCSYSE.2022.10056047

Article History:

Received:	11 June 2022
Last revised:	14 January 2023
Accepted:	14 January 2023
Published online:	19 May 2023

Reproduction of humanness based on eXtended intelligence: concept of artificial personality and its mechanism

Takayuki Fujimoto

Department of Information Sciences and Arts,
Toyo University,
2100 Kujirai, Kawagoe-shi, Saitama, Japan
Email: fujimoto@toyo.jp

Abstract: In this paper, we propose the concept of artificial personality (AP) and its mechanism, as a computer system in which users can feel human touch, or humanness. As computer performance improved in recent years, the capabilities of artificial intelligences (AIs) as well as the societal need for them grew, making AI the focus of public attention in every field. While current AIs are sophisticated, they are either categorised as industrial Narrow AI with limited uses or Toy AI intended for entertainment purposes. They are not general-purpose, versatile AIs with human touch, which can readily become a part of human life. There are a number of problems and challenges that need to be solved to realise versatile AI. One of them is an argument that questions whether the concept of AI or its method is appropriate for incorporating humanness in computer systems. Therefore, we discuss in this paper a method of reproducing humanness in computer systems based on the concept of AP rather than on that of current AI. We particularly focus on the idea of human and computer collaborating to form AP, and design a system according to the concept of eXtended Intelligence (XI).

Keywords: artificial intelligence; artificial personality; AP; eXtended intelligence; XI; artificial life; Chatbot; intelligent system; structure of personality data; date science.

Reference to this paper should be made as follows: Fujimoto, T. (2022) 'Reproduction of humanness based on eXtended intelligence: concept of artificial personality and its mechanism', *Int. J. Computational Systems Engineering*, Vol. 7, No. 1, pp.30–40.

Biographical notes: Takayuki Fujimoto is a Professor, Faculty of Information Sciences and Arts, Toyo University. In addition to he is a scholar, he is a Graphic Designer/Creative Director. He got his PhD from Yamagata University and got his Master degree of Knowledge Science from Japan Advanced Institute of Science and Technology (JAIST) and got his Bachelor of Education from Waseda University. His research is including the area of computer science and information design, cultural studies. He is a member of IEEE, and several others.

1 Introduction

1.1 Background

Since around 2006, a global phenomenon called 'The third artificial intelligence (AI) boom' has been observed. With a rapid advancement in computer performance, AI systems such as deep learning that need to solve complex equations are now widely used in everyday life (Fujimoto, 2018, 2023a, 2023b). In short, 'The Third AI Boom' means that the AI technology is now used for practical purposes in different aspects of our lives. The exponential development of AI has greatly changed the shape of digital life. Self-driving feature of cars and machine translation are now close to practical applications. AI-powered environments have brought more efficiency and automation, impacting our lifestyle. While 'AI' is now a well-recognised word, its idea is still not clearly understood by the public. This is because different AIs have different purposes and qualities depending on their uses. For example, most of the AI systems developed to a practical level today are Narrow artificial intelligences (Narrow AIs). They are dedicated programs for specific purposes, which are highly practical under specific situations but are not capable of achieving other purposes than the predefined one. Many possibilities of artificial general intelligence (AGI) that can handle varied situations and tasks are being explored. However, no AGI that can support practical or commercial uses is in sight.

1.2 AI and humanness

The top expectation for versatile AI is the ability to equip the computers with human-like behaviour. One example is creating computers that can perform human-level conversation. Since the early stages of AI research, developers have been experimenting with so-called chatbot, the computer program, which autonomously conducts automatic conversation. For instance, ELIZA is a famous chatbot developed by Joseph Weizenbaum in 1966. Many

different chatbots released in recent years have autonomous language learning ability based on machine learning and acquires language through conversation with human. Chatbots can provide minimal 'human-like' interaction to answer simple, typical questions or queries, using the words stored in the database and basic sentence structure (Arain et al., 2018). Chatbots are built with simple programs and can autonomously extend their own dictionaries, or knowledge, by learning new words and their usage through conversations (Kirsching et al., 1996; Hill et al., 2015). Though being a simple computer program, the chatbots give us far more human touch than Narrow AIs with advanced features. It is, so to speak, a system that autonomously forms its own personality through chatting with a large number of unspecified people.

For example, IKEA, a furniture retailer, has started using a chatbot system called Ask ANNA in 21 countries, which 20% of their customers utilise. It is now recognised as the brand mascot of IKEA while achieving a real success by lowering the annual growth rate of the number of call centre queries from 20% to 7%. Because Ask Anna is specialised in customer support for furniture-related queries, it is a simple, yet complete system, bringing convenient services. On the other hand, there are many challenges to implementing Conversational AI, which autonomously forms personality by machine learning. The biggest problem is that the auto-formed personality or knowledge can get overwritten in any way based on the trainer's/partner's personality and training. A personality that can be easily overwritten by a third-party user is at risk and can be dangerous, and people are not likely to find human-like qualities in such personalities. For example, in March 2016, Microsoft released Tay, an AI chatbot capable of autonomous natural language acquisition by machine learning. With 18 – to 24-year-old Americans as target dialogue partners, the language acquisition experiment was carried out, utilising Twitter. As a result, Tay's personality was shaped to repeatedly produce extreme, racist and immoral comments only 16 hours after its launch, causing Microsoft to stop the experiment immediately.

In January 2021, SCATTER LAB in Korea developed and launched Ilda, an AI chatbot, but it also learned to make sexist and racist comments, which forced the company to stop the service. Chatbot is a system, which uses machine learning to autonomously acquire language through chats with human and provide natural dialogue. Many types of simple chatbots have been developed to output combinations of statements stored in the dictionary by analysing the data of timely responses and conversational grammar. Although today's AI-powered, machine-learning chatbots may be different from their predecessors in terms of accuracy and functionality, they are essentially quite similar. Autonomous language learning means that the AI acquires knowledge from the conversation partner's speech and forms personality accordingly, even when the partner keeps making racist comments. At this stage, AI has no ability to discern good from evil or make conscientious, ethical judgments, where different interpretations are possible. It is difficult to form a personality that surpasses the equipped dictionary. Therefore, the following conditions are absolutely necessary for AI to properly acquire language and personality through autonomous learning: a conversation partner must be saintly; a person of integrity with strong morals; commonsensical and infallible; with universal ethics. Otherwise, the AI can automatically acquire the partners' bad habits, bias and immorality. There is a close relationship between personality and a speaker's choice of words. It is because one's personality is understood by others as expressed in his or her words. It would not be an exaggeration to say that language acquisition and use shape a person's personality (Freud, 1958).

Figure 1 (a) Yorishiro image by Kuniyoshi Utagawa (歌川国芳), 19th century and (b) sacred tree at Yuki Shrine (由岐神社) in Kyoto (see online version for colours)

(a)

(b)

1.3 Purpose of research

In this research, we propose a computer system that makes us feel human touch by defining the concept of artificial personality (AP) and designing its mechanism, instead of taking the previous approach based on AI technology (Aylett et al., 2017; Kadri, 2011, 2014; Sohn et al., 2012; Kanai and Fujimoto, 2018). While the idea of machine learning is to have computer systems acquire human-like intelligence through autonomous learning and realise humanness independently, the concept of AP is formed differently. In this research, we separate personality from intelligence and aim to create human-like behaviours using a computer program that simulates real person's personality. The proposed AP mechanism consists of three personality layers: character layer (CL), ghost layer (GL), and Skelton Layer (SL), into which real people's personality traits and patterns are extracted and categorised through observation of the subjects. Each of these personality layers is linked to the database based on each feature.

CL represents the public image of the person, or in other words, official personality.

GL represents the personality image, which is only recognised by the people close to the person such as relatives and best friends. This is generally expressed as the real nature of the person.

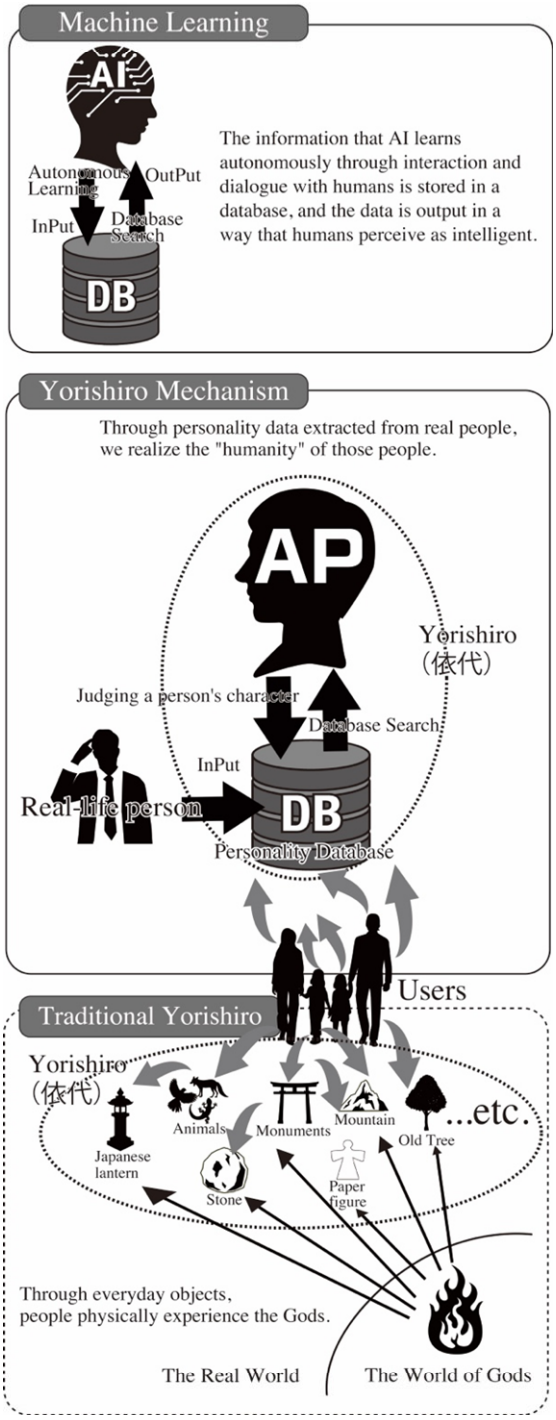
SL represents the image of universal personality, which all human beings have in common to some extent, and is not what applies to specific individuals. In this layer, no distinct individual personality or uniqueness is included.

By combining three personality layers, we design AP in which we see humanness. This mechanism is called Yorishiro Mechanism in this research. Yorishiro(依代) represents a Japanese idea toward gods and spirits, which has been passed down since olden days (Figure 1). It means the objects -ornaments, trees, houses, or places - in which gods or spirits dwell. Once they do, these objects are equated with the deities, and are worshipped. In short, the objects have invisible substance.

We assume that it is difficult for the 'unique intelligence and personality' – which AI autonomously acquires by machine learning – to provide humanness which people feel emotionally attached to. We also think that acquisition of humanness by computer system could be achieved by Yorishiro Mechanism, which is designed through elaborate observation and analysis of a real person as a model. This is because a computer-generated personality without Yorishiro is a fictional character, which is not linked to any actual person in real life. Figure 2 shows the difference between general machine learning and Yorishiro Mechanism regarding personality formation.

AP realised by Yorishiro mechanism would provide various application possibilities. For example, collecting from someone dying as much personality data on CL/GL and information as possible and storing them would enable the person to leave an 'automatic dialogue system that behaves like him/her'. With this, users can find his/her own personality in the system. It is, so to speak, 'permanent preservation of personality'. Unlike fictional character formation by machine learning, the system does not learn language or behaviour autonomously. Therefore, it never changes its personality through conversations or acquire knowledge that the model person could not gain after death. Only 'his/her personality' is technically kept and users can 'experience' talking with the deceased.

Figure 2 Personality formation by machine learning and Yorishiro mechanism



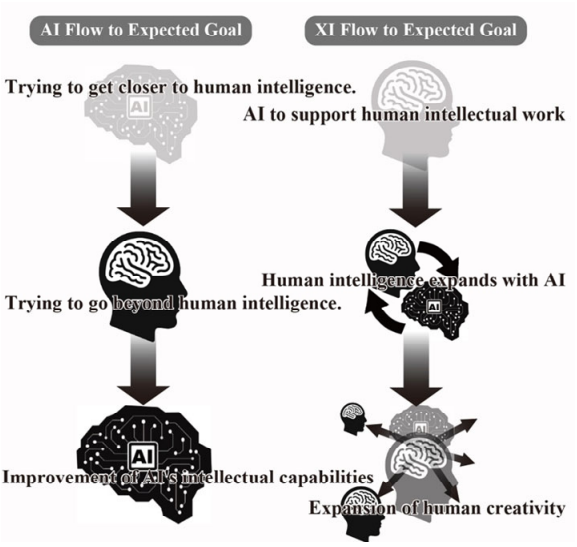
1.4 eXtended Intelligence as a post-AI concept

The term AI and its concept are now getting obsolete. As a matter of fact, IEEE and MIT media Lab have proposed 'eXtended intelligence (XI)' as an alternative concept and term to AI, which indicates a new form of computer intelligence (Ito, 2016, 2019). XI (also known as EI) is generally explained as follows.

XI is the use of AI to enhance human intelligence. EI blends the different strengths and weaknesses of the two types of intelligence. Human intelligence is used to direct the computational strength of AI through better free association and understanding. Human intelligence excels at processing sensory data, understanding, abstract thought and free association. AI excels at remembering, processing, prediction and analysis. While they differ, human intelligence and AI are good at balancing out each other's weaknesses and complementing each other's strengths. For example, where humans have difficulties with scale, computers can perform trillions of operations a second. Figure 3 shows the difference between traditional AI and XI (Karachalios et al., 2018).

Today, while AI outperforms human in specific intellectual tasks, it is also regarded as a competitor or a threat, which may deprive humans of their intellectual work. This is because a large part of such work is routinised. It does have aspects that require high intelligence, but not to the extent of utilising the unknown realms of human intelligence.

Figure 3 Comparison of AI and XI mechanisms



Examples of intellectual work include beating a human opponent in complex games such as chess and Shogi, or Japanese chess; negotiating over a certain topic for the best solution; performing bookkeeping and accounting tasks to plan a future budget; making predictions based on the past and present phenomena; and creating content which would gain popularity among the general public. Apparently, these are creative or intellectual human activities, but each of them actually is 'predictive work' based on search and analysis of big data. However, the types of intellectual work where AI can be useful are limited to those that 'appear' to be creative/intellectual human activities. Computers never behave in an unexpected or inexplicable manner, or irrespective of the past data, unless they are programmed to act ' randomly ' or 'crazy'. However, human creativity can be found in ways of thinking, ideas, and decision-making when they are unpredictable, uncalculated and spontaneous.

XI is based on the idea of computers that support and extend human intellectual activities instead of ones that are as capable as human and can replace human intelligence (Cabitza, 2020; Barack and Jaegle, 2022). It is not supposed to perform intellectual work in place of human or to compete against humans. This concept of XI suits various aspects of computing such as both current and ideal roles of computers, and the tasks computers are good at.

In the context of the XI concept, the master-servant relationship between human (master) and computer (servant) is never reversed (Fujimoto, 2023a, 2023b). Humans are actually required to control the computers by making decisions that are difficult for computers to make just by finding the best solution through a database search, on ethical views and moral values that every human has. In short, humans need to remain the 'ultimate decision-maker' with regard to ethics and morality to control the computers instead of attempting to teach them to computers. While the purpose of AI is to create a substitute for humans, XI is for limitless augmentation of human activities and society.

In this research, AP is to be shaped based on XI concept. Specifically, the computer will store, organise and utilise a large collection of human personality data by using its search function, huge memory and ability to perform high-speed calculation. On the other hand, instead of implementing an autonomous learning/development mechanism, humans are responsible for the final design as developers, and the computer serves as a database tool within the scope of this design. Personality of a particular person can be extracted only through observation by humans, and thus automatic extraction of data or patterns by machines is not feasible. Therefore, the proposed AP can be realised only through human-computer collaboration.

2 Relevant studies

Just as there are researches that overlap between Chatbots and versatile AIs, the studies on AP and the conversational AI research have much in common (Gunasekaran et al., 2013; Janzen, 2019; Gioti, 2020).

2.1 ALEXA (Amazon)

ALEXA, a virtual voice assistant system developed by Amazon.com, Inc., can properly provide audios such as music and radio as needed, in response to the user's voice commands. However, ALEXA is not a machine learning-based system with the function to autonomously learn to recognise the user's voice. It is well known that the final analysis of the commands is performed by, reportedly, thousands of dedicated staff members at Amazon since AI

alone cannot make good judgments in many cases. In short, the system can select a right output for the input but cannot autonomously judge what the input really means. Though the service appears to be versatile, in fact, it is difficult to call ALEXA an AGI.

2.2 Sophia (Hanson Robotics)

Sophia, developed by Hanson Robotics Limited, became a subject of news coverage when she was given citizenship of Saudi Arabia. However, she has only limited functions as a 'chatbot equipped with the ability to learn', and indeed her AI is far from AGI.

2.3 Geminoid (Hiroshi Ishiguro)

The first example that comes to our mind as research on giving computers a human touch is the real humanoid research by Hiroshi Ishiguro et al. The field of humanoid research, which aims to realise natural human facial expressions and movements by using realistic human mould, has seen a rapid progress in recent years with the advancement in technologies and materials. However, the field basically focuses on human-likeness of the mould and does not pursue humanness in terms of personality. To reproduce intangible aspects, extracting localised human functions, having computers reproduce them or using computers as substitutes for humans, may create limited humanness, but it is different from so-called humanness.

2.4 Singularity (Ray Kurzweil)

Ray Kurzweil states that, in 2045, "Singularity" will occur when the capabilities of AI and those of human brain reverse. However, a few contradictions and problems can be found in Kurzweil's idea. For example, he sees human brain as a high-performance calculating machine. It is unreasonable to compare computer and human brain in terms of their performance as calculating machines, because human brain's functionality does not lie in its calculation speed. Considering the rapid development in computational performance, it is certain that the current computers' processing power will surpass that of human brain in quite a short period of time. However, AIs cannot demonstrate something like aesthetic sense. This is because NO AI has the ability to judge the aesthetic values of an object. There is no basic research or set of specific guidelines that has been presented to make such judgments possible. Whereas humans perceive and appreciate beauty, AIs cannot do anything but perform database search and comparisons. Aesthetic value judgment is relative, but it also can be absolute. While 99% of a group of people feel a certain object is 'not beautiful', some may feel it is 'beautiful', and vice versa. Aesthetic values cannot be judged either by comparison or based on the historical data. There are also cases where people spontaneously and instantaneously judge whether something is 'beautiful' or 'not beautiful', or even change their previous judgment. Aesthetic judgments that humans make can be ambiguous, and sometimes people even find the ambiguity of the standard itself as 'beautiful'. How people go for a certain judgment is difficult to explain in scientific ways, and therefore, aesthetic value judgment by AI is impossible both theoretically and practically.

Relevant studies described in this chapter show how difficult it is for developers to realise AGI. Regarding Narrow AI, recent developments in AI technology have led to the most advanced forms of practical use. It is no exaggeration to say that, in today's society, we cannot live without the support of industrial AIs. On the other hand, we are far from practical application of AGI as its development is essentially not taking place, though AGI development has been considered ever since the birth of the AI concept in 1960s.

3 Consideration of personalities categorisation for artificial personality

3.1 Artificial personality by Yorishiro mechanism

AP by Yorishiro Mechanism is a different approach from autonomous intelligence generation by machine learning and computers. The aim of AP is to realise a computer system with which users can feel human touch, or humanness, by fusing Narrow AI and AGI. In other words, the proposed system is neither a new type of AI nor based on the concept of so-called next-generation AI. The approach of the proposed system completely differs from the concept or design philosophy of AI.

AP especially focuses on humanness brought by computer systems. Specifically, it aims to realise humanness by reproducing the qualities of a particular person. Personal qualities mean humanness, which everybody around this particular person clearly perceives. To implement humanness by computer systems, we think that AP needs to be created independently by separating personality from intelligence. AP is achieved by reproducing a personality modelled after a real person through the modelling process of meticulous observation, data collection and reproduction of the real person's personality. The proposed AP is different from the AI or intelligent behaviours of AI, which are generated autonomously and automatically by machines. In short, unlike AI, AP needs an actual person's personality as its model.

The biggest reason why AP needs a real person model is that there are a considerable number of cases in which humans behave irrationally or make unjustifiable decisions at times, which eventually results in the best solution that is so human and shows who they really are. Personalities or personal characters that make you perceive humanness cannot always be reproduced by computing.

The most distinctive characteristic of being human is illogical decision-making with inconsistencies, which could never happen in the course of database search, patterning and subsequent calculation. For AI, it is difficult to accept that such outcome is human, and this eventually leads to AI's meaningless output that has no human touch or

rationality inherent in computer. On the other hand, with the personality of a particular person model, AP makes the final decision based on 'what the model person would do.

3.2 Personalities and their categories

In this chapter, we describe the basic concept of AP, the system proposed in this research. Humanness that is reproducible with computers will be studied with a focus on personality instead of mould technology or partial intelligence.

Generally, personality is defined as follows:

- Personality: the type of person you are, shown by the way you behave, feel, and think. (*Cambridge dictionary*)
- Personality
  1 the quality or state of being a person
  2 the condition or fact of relating to a particular person
  3 the complex of characteristics that distinguishes an individual or a nation or group
  4 distinction or excellence of personal and social traits (MERRIAM-WEBSTER'S UNABRIDGED DICTIONARY)

To sum up, personality can be defined as a composite of elements and factors that characterise a person and are unique to him/her. In other words, the only way to identify an individual is the person's personality traits. Therefore, the following logic holds true.

1 Real person X exists. → personality A is generated

2 Personality $A \subset$ real person X

3 Personality A exists. → real person X is recognised

4 Personality $A \subset$ real person $X \rightarrow$ real person $X =$ real person X is recognised

5 Personality $A \coloneqq$ real person X

Personality A, which characterises person X and with which you recognise person X, has nothing to do with academic achievement or information processing capability. Instead, in many cases, humanness or personality unique to the person can be found in unanticipated, spontaneous and unrelated actions or actions that are unpredictable from the past data. It is difficult for users to perceive humanness in the personality program that is autonomously generated by computer's machine learning. This is based on the same reasoning that improving processing power of the computer system does not reproduce human brain functions.

We think that AP, in which users perceive humanness, can only be achieved by reproducing a real person's personality, instead of by creating the unique personality autonomously generated by machine learning. Therefore, without an actual model Person X and personality A, *AP* would not be made possible.

When a real person X exists, the person possesses a unique personality A. In this paper, we suppose that the unique personality A is comprised of *personality-a(Pa), personality-b(Pb), and personality-c(Pc)*.

- Unique personality A
  - Pa (Personality-a): personality understood/grasped by oneself
  - Pb (Personality-b): personality that is not understood/grasped by oneself
  - Pc (Personality-c): personality recognised/grasped by a third party and society

3.3 Personality understood / grasped by oneself: Pa (Personality-a)

Everyone has an understanding of yourself, or of who you really are, apart from others' understanding of you. However, your personality as you understand it is hard to prove that it is your true self, using objective indicators. This is because your understanding of your own personality is nothing but an 'assumption', just like a third person's understanding of your personality is. The unique personality A, which person X possesses, represents person X.

Assuming $A \coloneqq X$, *Pa* (personality understood by oneself) belongs to personality A as well as to person X. However, this does not prove that personality A is exactly the same as *Pa*. This is described in symbols as follows:

$$Pa \subset (A \coloneqq X) \nvdash (A :\Leftrightarrow Pa)$$
$$\therefore Pa \models \neg A$$

3.4 Personality that is not understood/grasped by oneself: Pb (Personality-b)

Everyone has *personality-b* (*Pb*), which could not be understood or grasped by oneself. *Pb* constitutes Personality X which characterises person X, but *Pb* does not represent the entire personality A. Assuming $A \coloneqq X$, *Pb* characterises Person X but is not proved to be exactly the same as Personality A. Therefore, Pa is not the same as *Pb*, either, and *Pa* and *Pb* are different personality elements.

This is described in symbols as follows:

$$Pb \subset (A \bigvee X) \nvdash (A :\Leftrightarrow Pb)$$
$$Pa \subset (A \bigvee X) \nvdash Pb \subset (A \bigvee X)$$
$$\therefore (Pa \neq Pb) \models \neg A$$

3.5 Personality recognised/grasped by a third party and society: Pc (Personality-c)

Assuming $A \coloneqq X$, everyone has *personality-c* (*Pc)*, which is recognised by the society and people around oneself, apart from *personality-a* (*Pa*) understood by oneself. *Pc* is a personality element that characterises person X as viewed by people. As in the cases with *Pa* and *Pb*, *Pc* does not represent the entire personality A. Also, *Pc* is not proved to be same as personality A and is a different personality from *Pa* or *Pb*.

This is described in symbols as follows:

$$Pc \subset (A \vee X) \nvdash (A :\Leftrightarrow Pc)$$
$$Pa \subset (A \vee X) \nvdash Pc \subset (A \vee X)$$
$$\therefore (Pa \neq Pb \neq Pc) \models \neg A$$

Based on the above structure, personality *A*, which is comprised of three personality elements, characterises person *X* and makes it possible to identify person *X*. On the other hand, each of the three personality elements has different attributes and values, and these personality elements cannot objectively prove that personality *A* is true. Therefore, a person's 'personality' that people can recognise and understand is defined as *Pc*, and the proposed AP means the reproduced *Pc*.

In this study, we intend to reproduce the unique personality of an individual, which is equivalent to humanness, by defining and simulating *Pc* as the personality which, based on an objective view, characterises person *X*. Even if person *X* saw *Pc* as unacceptable, it would not change his/her external personality, which characterises person *X*. For example, if a serial killer, who is believed to be a heinous criminal not only by lawyers and judges but also a majority of the public, insisted that he/she is not the kind of person as reported, the society's shared perception of the killer would never change. In this case, even if the killer actually were a person of excellent character, that personality would never be known to or believed by the public, and he/she would only be identified as 'a heinous serial killer'. In short, AP *Y* is the reproduction of actual person's external personality recognised by people. When people who know person *X* see AP's following actions, they can find person *X*'s uniqueness, or humanness, in AP *Y*: response to the partner's comment; reaction to the partner when he/she does not respond; or any actions unique to person *X*.

4 Outline of artificial personality

4.1 Modelling of artificial personality

Research on human personality is being performed in various fields such as psychology, cognitive Science, Life Science, and intelligent informatics. Human personality is not one-dimensional but so complicated that it is extremely difficult to represent with a simple model.

Research on human personality is being performed in various fields such as psychology, cognitive Science, life Science, and intelligent informatics. Human personality is not one-dimensional but so complicated that it is extremely difficult to represent with a simple model. In this study, we design a model that is more focused on human touch or humanness based on the prior human personality studies. We think of the AP design model as a nested structure of personality with three layers: *CL, GL,* and *SL*.

A major clue to the composition of the human personality comes from Sigmund Freud's work. According to Freud's theory 'Topograph', human personality is considered to be composed of three layers: conscious, preconscious, and unconscious.

- Conscious: mental processes that one is aware of oneself
- Preconscious: a mental process that one recalls if one is aware of it
- Unconscious: psychological processes that one is not aware of or cannot be aware of

On the other hand, when we construct an AP as a computer system, we cannot use Freud's theory to analyse the human personality. This is because it is difficult to design or program the unconscious as a computer system.

In this study, we design a model that is more focused on human touch or humanness based on the prior human personality studies. We think of the AP design model as a nested structure of personality with three layers: *CL, GL,* and *SL*.

4.2 CL: character layer

CL is the external part of personality. It is a layer that is outlining the personality. Specifically, this part is recognised by outsiders as the kind of person you are. What is notable about this layer is that it is your personality as seen by outsiders, not the personality you believe you possess or your true personality.

Here are the examples:

- Real person X_1 is a talkative and unique comedian.
- Real person X_2 is an aggressive and speculative businessperson.

Generally, we judge people's personality by their visible characteristics. Therefore, *CL*, which is at the surface of the personality, defines the person, leaving the outsiders an impression of the person. However, as a matter of course, the layer does not necessarily mean the person's true personality. In some cases, people may be acting to hide their true personality that they think they possess or be changing their personality against their will to perform work efficiently. Entertainers and celebrities are often thought to have a certain personality according to their external character traits, but if you had a chance to meet them in person, not a few of them would give you an impression that do not correspond to their external personality traits *CL*

4.3 GL: ghost layer

GL represents the real nature or real personality of a person, which *CL* does not cover. *GL* is the personal layer that represents the personality, which you come to know through a close, personal relationship or through living with the subject person.

Below are some of the examples.

- Real person X_1 is famous as a talkative and unique comedian but is quiet and serious offstage, speaking very little.

In this case, talkative and unique fall under *CL*, and quiet, speaks little and serious are classified into *GL*.

- Real person X_2 is an aggressive and speculative businessperson at work but at home, he is a mild, loving father.

In this case, aggressive and speculative are classified into CL, and mild and loving into GL. GL is further divided into the elements shown in Table 1.

Table 1 Elements that constitute the ghost layer

Ghost layer: GL
GLx1: personality only known to him/herself
GLx2: personality known to close relatives such as family members
GLx3: personality known to the people who has a relationship with the person such as friends
GLx1 ⊆ GLx 2 ⊆ GLx 3 ≒ character layer

Table 2 Basic response patterns that Skeleton layer provides

Skeleton layer		
INPUT	➡	OUTPUT
<Attitude>		
• aggressive	➡	• discomfort, antipathy
• moderate	➡	• receptive, agreement
• support	➡	• pleasure, appreciation
• no interaction	➡	• loneliness
• too much interaction	➡	• annoyance
<Information>		
• not enough	➡	• anxiety, distrust
• moderate	➡	• security, confidence
• superfluous	➡	• chaos, difficulty
<Correspondence>		
• danger	➡	• Avoid
• hardship	➡	• Avoid
• safety	➡	• Secure
• comfort	➡	• Secure

4.4 SL: Skeleton layer

SL is the layer for universal personality, which all human beings have in common to some extent. Unlike *CL*, personality in this layer is not for the third-party people to identify a person with. For example, the personalities shown below are commonly seen in almost all the humans. The personality traits in the *SL* are seen in responses to social interactions in three basic patterns: negative, positive and neutral.

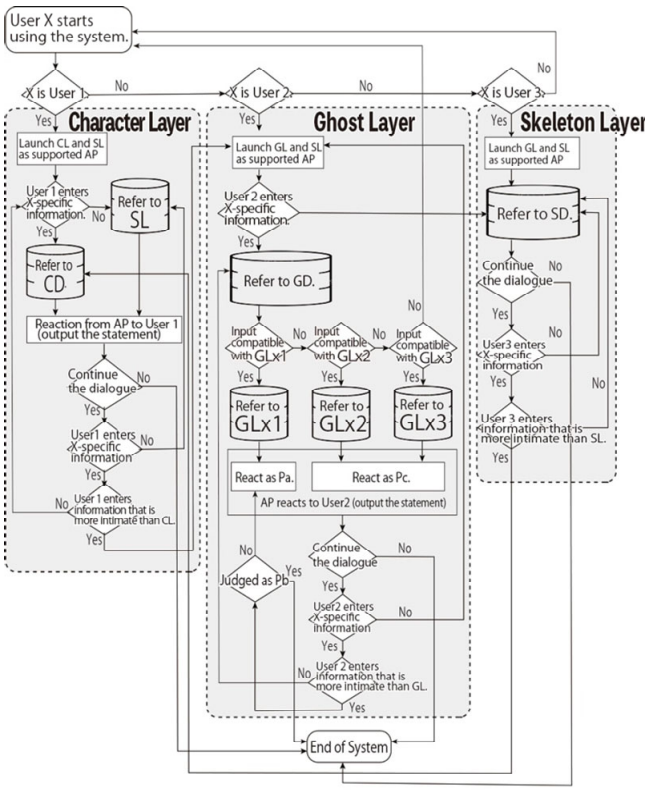
Here are the examples.

- All humans have negative emotions when somebody gets angry at them. (example: pains →aversion)
- All humans have positive emotions when praised. (example: delicious food → happiness)
- All humans judge when they encounter something unknown or something they have never been interested. (example: no interest → ignore / become interested → positive/not interested → negative)

Skeleton layer consists of these universal patterns of basic responses to social interactions. These basic patterns are listed in Table 2.

With regards to SL, humans respond with one of the basic responses – negative/positive responses or judgments, depending on the type of external Input. These basic responses are commonly seen among almost all humans. When people 'have negative emotions as somebody gets angry at them', in many cases they get angry, feel sad, cry, lose their temper, fight back or get lost. The response differs depending on each cause, but all the responses are basically negative. More detailed and specific responses that characterise the person belong to *CL*, the layer above *SL*.

Figure 4 Flow chart of AP algorithm



User1: User attributes corresponding to CL
User2: User attributes corresponding to GL
User3: User attributes corresponding to SL

Pa(**Personality-a**): personality understood/grasped by oneself
Pb(**Personality-b**): personality that is not understood / grasped by oneself
Pc(**Personality-c**): personality recognized / grasped by a third party and society

GLx1: personality only known to him/herself
GLx2: personality known to close relatives such as family members
GLx3: personality known to the people who has a relationship with the person such as friends

GLx1⊆GLx 2⊆GLx 3≒Character Layer

5 Algorithm for AP

5.1 User attribute

Figure 4 shows the algorithm for the text-based conversation AP.

At first, user *X* (conversation partner *X*) needs to select his/her attribute from U1, U2 or U3, which corresponds to each of AP's three layers. By selecting one of the three attributes below, the corresponding personality layer is activated.

- User1: user attributes corresponding to *CL* which generally apply to users who know Real Person *X*
- User2: user attributes corresponding to *GL* which apply to close friends and relatives who know real person *X* well
- User3: user attributes corresponding to *SL* which apply to the people whose personality has no individuality or uniqueness. They give typical and universal responses.

This system refers to AP databases for the three personality layers and generates responses according to the user attributes. Also, if the conversation partner's comments do not correspond to the user attributes preset for the conversation, the system will decide to refer to the databases of other user attributes. *CL*, *GL*, and *SL* personality layers refer to three databases, *character database* (*CD*), *ghost database* (*GD*) and *Skeleton database* (*SD*), which store corresponding comments and behaviours for each attribute.

5.2 Three-layer model for AP database

In this research, AP database is divided into two categories: words and deeds, and three types of databases that correspond to three AP personality layers are constructed:

- CD
- GD
- SD

Of the three databases above, SD stores data for universal personality, which is barely affected by individual's uniqueness, sense or character, in light of the human cognitive mechanism. On the other hand, CD and GD are data that characterise or identify the subject person with.

5.2.1 Character database

CD is personality data that consists of the person perception that people around the subject person, society or the community has about him or her. Therefore, to build *CD*, we collect the perception data about the subject person. Specifically, the database is built using the following methods.

1 Questionnaire survey for the people, society and community around the subject
2 Observational research by developers and hearings with the subject

Therefore, *CD* is easy to construct for famous people because the impressions and perceptions of them have been widely shared among the public. Typical items to be collected for *CD* are listed in Table 3.

Table 3 Words/deeds data to be stored in character database

Words	• Real person *X*'s favourite phrase
	• Things real person *X* would say
	• Responses real person *X* is likely to give
	• Stories and memories about real person *X*
	• Real person *X*'s expertise
Deeds	• Timing and speed at which real person *X* utters
	• Real person *X*'s habits and traits that reveal in conversation
	• Timing and speed at which real person *X* responds

5.2.2 Ghost database

GD is personality data shared among people who have a close relationship with real person *X* such as relatives and best friends. As a matter of course, *CD* and *GD* have different data content. In some cases, *GD* and *CD* can be quite the opposite in content. *GD* is thus built on the results of the questionnaire for and hearings with the relatives and close friends. Survey questions are the same as those for *CD*, with some additional questions on the following items.

Examples of main items to be collected for *GD* are shown in Table 4.

Table 4 Words/deeds data to be stored in ghost database

Words	• What real person *X* says only to people in close relationship with him/her
	• Secrets which real person *X* tells only to these people
Deeds	• Real person *X*'s unique responses in interactions with these people
	• Real person *X*'s behaviour which is different from his/her behaviours known to the public

5.2.3 Skeleton database

SD is basic and universal personality data that all humans have in common. *CD* and *GD* may produce different responses (outputs) to inputs, or stimuli, depending on each person's uniqueness or nature. However, it is not uncommon that people react in a similar way. As an example of SD response, to an input from an external source that conveys praise, a basic output expressing 'happiness' is produced. Though *CD* and *GD* may produce distinctive outputs such as downplaying the compliment to hide embarrassment, or showing humility, but at the base of the different outputs is the feeling of happiness. Naturally,

everybody feels happy when they are praised, and although there may be differences in the way they express how they felt, the feeling itself is universal.

Because data in *SD* is not intended for identification or characterisation of a person or personality, *SD* does not store specific language data (comments data) and preserves only behavioural data.

SD is constructed based on the research of general cognitive science. Main items to be collected for SD is shown in Table 5.

Table 5 Behavioural data to be stored in Skeleton database (SD)

Deeds	• Positive output to positive input • Negative output to negative input • Output 'judgment' to the input in which the AP does not have any knowledge or interests

6 Conclusions and future work

In this paper, we proposed a concept of AP and its mechanism as a computer system to represent human touch or humanness and modelled the structure. We described that giving humanness to computer systems is achievable only by simulating a real person, not by using the machine learning-based, autonomously formed personality. In other words, to create human touch using computer programs, a real person's personality on which AP is to be based on is absolutely necessary. This mechanism is called Yorishiro mechanism in this research.

On the premise that an AGI in which humans can perceive humanness will never be realised with the use of recent, eye-catching AI concepts and its developmental methods, we proposed an approach based on human-computer collaboration and the existence of a real person model in order to realise an AP with humanness. We proposed the concept of AP based on Yorishiro mechanism and developed a theoretical model of its structure. As our future work, we will go forward with practical implementation of AP, utilising the AP concept and Yorishiro mechanism.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP22K12134.

References

Arain, A., Manzoor, A., Brohi, K., Haseeb, K., Halepoto, I.A. and Korejo, I.A. (2018) 'Artificial intelligence mark-up language based written and spoken academic chatbots using natural language processing', *Sindh University Research Journal (Science Series)*, Vol. 50, No. 1.

Aylett, M.P., Vinciarelli, A. and Wester, M. (2017) 'Speech synthesis for the generation of artificial personality', *IEEE Transactions on Affective Computing*, Vol. 11, No. 2, pp.361–372, IEEE.

Barack, D.L and Jaegle, A. (2022) *Extended Intelligence*, arXiv preprint arXiv:2209.07449, https://doi.org/10.48550/arXiv.2209.07449.

Cabitza, F. (2020) *From Artificial Intelligence to Humanistic Intelligence and then Extended Intelligence* [online] http://sigchitaly.eu/wp-content/uploads/2020/07/Cabitza_HCI4AI_Syllabus_paper_8.pdf.

Freud, S. (1958) 'A note on the unconscious in psycho-analysis', in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, *The Case of Schreber, Papers on Technique and Other Works*, Vol. 12, Nos. 1911–1913, pp.255–266.

Fujimoto, T. (2018) 'Ideology of AoD: analog on digital-operating digitized objects and experiences with analog-like approach', *1st International Conference on Interaction Design and Digital Creation Computing*, pp.901–906.

Fujimoto, T. (2023a) 'NEOTEN¥ (NExt optimum trade environment of ¥en): a cryptocurrency with market capitalization linked to the Japanese Yen, non-speculative crypto-assets with economic security', *Applied Systemic Studies*, pp.53–63, Springer.

Fujimoto, T. (2023b) 'Potential of eXtended intelligence (XI) for extending human expression from digitization of analog elements', *Applied Systemic Studies*, pp.31–39, Springer.

Gioti, A-M. (2020) 'From artificial to extended intelligence in music composition', *Organised Sound*, Cambridge University Press, Vol. 25, No. 1, pp.25–32.

Gunasekaran, S.S., Mostafa, S.A. and Ahmad, M.S. (2013) 'Personal and extended intelligence in collective emergence', *13th International Conference on Intellient Systems Design and Applications.*

Hill, J., Randolph Ford, W. and Farreras, I.G. (2015) 'Real conversations with artificial intelligence: a comparison between human–human online conversations and human chatbot conversations', *Computers in Human*, ELSEVIE, Vol. 49, pp.245–250.

Ito, J. (2016) *Extended Intelligence*, Joi Ito's PubPub [online] https://assets.pubpub.org/03x0w4r8/f875537b-6bfa-4ec2-8579-2e90214f672a.pdf.

Ito, J. (2019) *Forget about Artificial Intelligence, Extended Intelligence is the Future* [online] https://fully-human.org/wp-content/uploads/2019/07/Forget-about-artificial-intelligence-extended-intelligence-is-the-future.pdf.

Janzen, R. (2019) *Extended Intelligence and Mediated Reality Veillametrics in the Space, Time, Frequency, and Amplitude Domains*, ProQuest Dissertations Publishing.

Kadri, F.L. (2011) *The Design and Validation of an Artificial Personality*, Kybernetes, Vol. 40, Nos. 7–8, pp.1078–1089.

Kadri, F.L. (2014) 'Understanding and learning to reconcile differences between disciplines through constructing an artificial personality', *Kybernetes*, Vol. 43, Nos. 9–10, pp.1338–1345

Kanai, Y. and Fujimoto, T. (2018) 'Proposal and development of artificial personality (AP) application using the ''requesting'' mechanism', *Computational Science/Intelligence and Applied Informatics (Studies in Computational Intelligence)*, pp.13–24, Springer.

Karachalios, K. and Ito, J. (2018) 'Human intelligence and autonomy in the era of 'extended intelligence', *The Council on Extended Intelligence* [online] https://globalcxi.org/wp-content/uploads/CXI_Essay.pdf.

Kirsching, I., Koyama, M., Aoe, J-I. and Tomabechi, H. (1996) 'The time-sliced paradigm–a connectionist method for continuous speech recognition', *Information Sciences: an International Journal*, Vol. 93, Nos. 1–2, ELSEVIER, pp.133–158, https://doi.org/10.1016/0020-0255(96)00083-7.

Sohn, K., Krishnamoorthy, S., Paul, O. and Lewis, M.A. (2012) 'Giving robots a flexible persona: the five factor model of artificial personality in action', *12th International Conference on Control, Automation and Systems.*