Journal on Interactive Systems, 2025, 16:1, doi: 10.5753/jis.2025.5465

Game-theoretic Methods for Differentiating Systems with Artificial and Anthropogenic Intelligence

Pavel Konyukhovskiy [**Herzen State Pedagogical University** | ***kon_pv@mail.ru***]

Institute of Economics and Management, Department of Industrial Economics and Finance, 2 Chernyakhovsky str., St. Petersburg, 191186, Russia.

Received: 22 January 2025 • **Accepted:** 16 May 2025 • **Published:** 01 June 2025

Abstract: The **aim** of the article is to develop procedures for differentiation (identifying) types of intelligent systems at the conceptual level. It is assumed that such systems can be built using both anthropogenic (natural) and artificial intelligence. The relevance of the goal is determined by the spread in the modern world of technologies called "artificial intelligence", which have a radical all-round impact. The research is based on the **methods** of strategic game theory. The main hypothesis of the study is reduced to the assumption that different types of intellects in game situations will correspond to stable characteristic models of strategic behavior, by which their differentiation can be carried out. The paper consecutively considers differentiation procedures based on bimatrix games of the class "family dispute" (games with two Nash equilibrium in pure and one in mixed strategies), differentiation procedures based on static Bayesian games, differentiation procedures based on signaling games. The differentiation methods are based on the hypothesis that carriers of different types of intelligence will consistently (significantly) differ in their strategic behavior in the game. In the conclusion the principles of comparative analysis of different procedures are formulated. The main **result** of the study is a set of differentiation procedures based on game-theoretic models. These procedures are free from the disadvantages inherent in expert approaches.

Keywords: Game-theoretic Methods, Artificial Intelligence, Anthropogenic intelligence, Differentiation of Intelligent Systems, Strategy Games, Bayesian Games, Mechanism Design

1 Introduction

The development of technologies united under the general heading of "artificial intelligence" (AI) has become one of the global fundamental trends of recent years. Despite the fact that the concept itself as such entered professional and journalistic discourse literally from the moment the first computers appeared, it is currently acquiring qualitatively new meanings and senses.

Separately, it is necessary to pay attention to the vagueness of approaches to defining the concept of AI itself. It is often used as a marketing brand to attract the attention of potential consumers to a product or technology promoted by a particular manufacturing company. In advertising for a significant number of modern gadgets (smartphones, tablets, etc.), it is invariably emphasized that they operate using artificial intelligence.

In any case, it is impossible to ignore the fundamental question of "where is the line separating complex, advanced software from artificial intelligence?" Often, "intelligence" is taken to mean the incomprehensibility (non-obviousness) of the logic of a particular program for its user.

The fundamental position of this paper is that at the present moment there is no reason to talk about artificial intelligence in the strict sense of the term "intelligence". The concept of intelligence is closely related to such categories as "personality" and "consciousness". Intelligence in the full-scale interpretation of this concept is characterized by the presence of individual independent subjectivity and, fundamentally, an individual system of goals and interests, which presupposes, among other things, the solution of problems of ensuring the sustainability of its life cycle. Here, comparisons with the intelligence of living organisms are quite appropriate, which, regardless of the degree of subordination of their position in the anthropocentric world, have as their initial priority goal-setting with their own survival (life support).

The classification of intelligence into natural (anthropogenic) and artificial cannot be considered as the only alternative. There are other approaches to constructing typologies of intelligence and intelligent systems. In particular, one can mention Gardner's theory of multiple intelligences, Gardner [2011].

At the moment, what is widely and loudly called artificial intelligence are programs that implement the functions of input and primary processing of information in formats traditional for the human world. First of all, as a rule, we are talking about the recognition of audio-visual information, speech, recognition of commands and requests given in the form of traditional human speech. At the same time, we are not talking about the fact that "intelligent" programs implement any of their own goals. They are still strictly focused on the goals of their creators, owners and users. It is more correct to ask the question about the effect of the "illusion of intelligence" that occurs when a person communicates with a software and hardware complex. As it becomes more complex, more and more people have the impression of reasonable behavior on the part of this complex. To a large extent, this is due to the gap in ideas about the capabilities of technical means. The effect of communication with an "intelligent partner" arises as a reaction to qualitatively new capabilities of programs that look like something unusual and unfamiliar against the background of experience and practice of past periods. It is likely that as public opinion becomes accustomed

to the new situation, this effect will weaken.

In light of the above, it is advisable to differentiate between concepts and, accordingly, to separate artificial intelligence in the strict sense (that is, something that does not currently exist), on the one hand, and artificial intelligence in the broad sense of the word, on the other. The latter can be understood as advanced software tools that implement neural network technologies, large language models, complex intuitive image recognition systems, etc. Perhaps, for artificial intelligence in its current state, the names quasi-intelligence or proto-intelligence would be more appropriate.

Despite the limitations of "artificial proto-intelligence", it is impossible to deny its influence on the socio-economic and socio-political processes of the modern world. The following areas can be cited as the most striking and illustrative examples.

There is almost a consensus that AI technologies will radically reformat the labor market in the near future, leading to the disappearance of a significant number of professions. A "cascade effect" is reasonably expected: the reduction of professions in some spheres will lead to their disappearance in related spheres. Expectations associated with a drop in the standard of living of the strata that have lost employment and a subsequent decrease in solvent demand seem quite rational against the background of these hypotheses. Of course, it is impossible to rigidly postulate the inevitability of "classical crises of overproduction according to Marx." However, the possibility of such scenarios also cannot be denied.

The problem of using AI in public administration is extremely important and at the same time controversial. One cannot help but notice the subtlety and blurriness of the line separating the technical aspects of management activities from decisions that are of a political nature. Access to global databases in combination with powerful tools for their analysis in the modern world is becoming the most important condition for maintaining political power. For example, the concept of a "smart city" has become quite popular in the modern world. At the same time, it is impossible not to recognize the existence of objective contradictions and inconsistencies in the interests of various segments of the population. As a result, we are faced with the question "based on what criteria of their priorities will the mind that manages a smart city resolve these contradictions?"

The spread of AI technologies in the field of education actualizes such problems as the effectiveness of AI as a tutor, medium-term and long-term consequences of AI pedagogical activity, the dilemma of choosing "teach a person" vs. "spend money on developing and implementing a robot", the specifics of functioning and new patterns of functioning of a mixed human-machine educational environment, forms of deception, corruption, substitution of goals in this sphere.

There is no doubt about the importance, relevance and prospects of using AI in healthcare and medicine, where they play the role of the most important resource for improving the quality of life. At the same time, it is impossible to ignore the problems associated with inequality of access to services, responsibility, privacy, trust in radically new forms of services.

Discussions about the use of AI in political and armed conflicts are particularly acute. Again, the line separating AI as a tool in the hands of decision-makers from AI as an actor in conflicts, independently making decisions with a not entirely clear system of motivations and goals, is extremely thin and opaque. In recent years, a growing number of works have appeared devoted to these aspects of the development of the modern world, see, for example, [Yakovleva *et al*., 2025].

Taking into account the listed items, we must state the importance of the tasks of modeling the relationship between natural (anthropogenic) intelligence (NI) and artificial intelligence (AI). In particular, methods for solving the problems of identifying the type of intelligence with which interaction is carried out, or, which is more terminologically correct, the tasks of differentiating (segregation, separation) types of intelligence, are becoming relevant and in demand.

The term "differentiation" is quite deliberately included in the title of this paper. The decision to use it was made to draw a distinction between classical classification and clustering methods and the approaches developed in this research.

The structural organization of the paper is based on the principle of successive complication of the considered game-theoretic models. At the initial stage, differentiation procedures based on the simplest classes of strategic games (bimatrix games) are considered. Then, a transition to static games with incomplete information (the so-called static Bayesian games) is carried out. Their use makes it possible to implement the provisions of the mechanism design theory on separation procedures. As the next step, a transition to an important particular step of dynamic games with incomplete information is considered. On their basis, algorithms for separation of types of intelligent systems operating in the "signal-reaction" mode can be constructed. In conclusion, possible methods of comparative analysis and evaluation of various methods for solving the problem of differentiation of types of intelligent systems are proposed.

2 Previous Studies and Related Issues

The problems of differentiation (separation) of systems with artificial and natural intelligence currently have a fairly rich history.

In this regard, it is necessary to mention first of all the well-known article by Turing [1950], in which the famous Turing test was formulated. The test is aimed at identifying the capabilities of a machine (computer, robot) to impersonate a human. In the most famous formulation of the Turing test, the judge has the opportunity to conduct a conversation with both a person (people) and computers (machines). Accordingly, a situation in which the judge is unable to clearly distinguish a machine from a person means that the machine has passed the Turing test.

The Turing test is often associated with the search for an answer to the question "can a machine think?" At the same time, many authors note the weaknesses of such a "philosophical" formulation of the problem. The possibilities of its constructive resolution are significantly limited by the vagueness and contextual dependence of the concepts of "intelligence" and "consciousness", as well as the ambiguity of the meanings invested in the concept of "thinking". Ultimately, it is very, very difficult to outline the line separating thinking from the routine execution of an algorithm. Especially if

the algorithm is quite complex (contains a complex system of conditions).

The general context of the test, in which the computer is supposed to "fool" a human, has also come under criticism. Such a result can be achieved through a machine's poorly understood imitation of human behavior, which does not mean that it is implementing complex thought processes.

Since this article is being written in the realities of 2024, it is impossible to ignore the surge of publications dedicated to Chat GPT-4 passing the Turing test, see, for example, [CNews, 2023]. It is symptomatic that the headlines of such articles, intended for a mass audience, are mutually diametrically opposed: from "Chat GPT passed the Turing test" to "Chat GPT failed the Turing test".

In the general flow of works on game theory topics, a direction has been formed related to the experimental verification of theoretical concepts. Groups of students studying courses with game theory content act as a natural and constructive base. For example, in [Dixit and Skeath, 1999] an idea is put forward suggesting the use of simple games with bargaining at the beginning of the course. Research experiments aimed at identifying students' behavioral choices when playing classical models of strategic games, such as the prisoner's dilemma, are quite popular and fruitful. In this case, the subject of study is the tendency (or, conversely, the reluctance to cooperate), i.e. the proportion of those who choose the Nash equilibrium and those who play a solidary, more advantageous, but unstable situation. Here, one cannot help but pay attention to the fact that a logical continuation of such studies is their extension to mixed groups consisting of carriers of anthropogenic and artificial intelligence.

In a series of studies devoted to the application of game-theoretic analysis methods to the comparative analysis of human and artificial intelligence behavior models, it is necessary to mention the article [Ghasemi *et al*., 2020]. The results of a series of drawings of the game "rock-paper-scissors" are considered, in which the opponent of a person (anthropogenic intelligence) is a neural network. In this case, a series of training and real games are distinguished. A differentiation analysis of anthropogenic participants by gender is separately conducted. According to the statistics of the results published by the authors, the neural network wins (in approximately 60% of games).

The paper [Matsumoto *et al*., 2012] examines the issues of classifying the logic of reasoning of participants in the game "rock-paper-scissors". In particular, the classification of "automatic judgment". The focus of the consideration is on the "c-means" methods.

Among the works devoted to the issues of classifying the actions of players in rock-paper-scissors games, the article [Gang *et al*., 2017] should be highlighted. The authors examined methods of classifying signals using a multilayer perceptron.

To one degree or another, the problem of using rock-paper-scissors games in conjunction with artificial intelligence systems is addressed in articles [Cenggoro *et al*., 2014], [Ali *et al*., 2000], [Hasuda *et al*., 2007], [Hu *et al*., 2019], [Salvetti *et al*., 2007].

It is worth paying attention to the studies conducted in related fields. The methods and approaches implemented within the framework of the LMSYS.ORG project [Lianmin *et al*., 2023] are quite interesting. In particular, Chatbot Arena is a platform for testing large language models (LLMs) functioning in the anonymous crowdsourcing mode. Participants who actively join the project become judges of the competition. In each round, two anonymous bots are compared. Based on the results, pairwise ratings of bot programs are calculated and regularly updated. The Elo coefficient, traditional for games such as chess, checkers or go, is used to compare the "players".

The problems of correlation between positive and negative consequences of intelligent information and digital technologies have become one of the leading trends in modern scientific literature. In terms of the significance of these problems, such studies as [Mozikov *et al*., 2024] are of interest. The authors consider the issues of security and correspondence of large language models (LLM) to human behavior. The paper actively applies the apparatus of strategic repeated games and bargaining games to study the emotional impact on ethics and decision making based on LLMs. In particular, the paper substantiates the thesis that LLMs are subject to emotional biases that are influenced by model size, matching strategies, and primary pre-training language.

No less complex and significant are the problems of consequences of extensive spread of new technologies (LLM, neural networks, GPT, etc.) in the sphere of education. Works [Garkusha and Gorodova, 2023], [Ivakhnenko and Nikolskiy, 2023] are devoted to the analysis of both objective advantages and potential threats of this process. We cannot but recognize that the dual nature of new software tools is a weighty argument in favor of the relevance of differentiation tasks. The procedures of introducing new intellectual tools and means are characterized by exceptional multidimensionality. In particular, a non-transparent set of moral problems is generated. In essence, the processes of transformation of classical university ethics are underway. At the same time, a set of requests is being generated in terms of reforming the methods of evaluating the results of the educational process. This, in turn, may lead to a rethinking of the criteria for the effectiveness of the functioning of universities in general. The works [Konyukhovskiy, 2022], [Kolyshkin *et al*., 2023] can be named as an example of research on these issues.

The idea of using strategic games to build differentiation procedures for identifying artificial intelligence seems quite natural, if not obvious. At the same time, one cannot ignore the equally fruitfulness of approaches involving the application of cooperative games. This direction can be considered as one of the possible perspectives for the implementation of the present study. Cooperative game-theoretic models, which have successfully demonstrated their efficiency in describing the regularities of interaction of the world centers of power [Konyukhovskiy and Malova, 2013], [Konyukhovskiy and Malova, 2015], [Konyukhovskiy and Kholodkova, 2015], with a high probability can be successfully extended to situations of coalition behavior of players (agents), who are carriers of different types of intelligence.

Table 1. Basic game.

		Player II	
		L	**R**
Player I	**U**	a_{11} | b_{11}	0 | 0
	D	0 | a_{22}	b_{22}

3 Basic Game-theoretic Model of Differentiation of Intelligent Systems

The prospects of potential contradictions and conflicts between natural and artificial intelligent systems undoubtedly increase the importance and relevance of game-theoretical methods in relation to differentiation problems. The main hypothesis of the game-theoretic approach is the assumption that there will be stable differences in the strategic choices made by anthropogenic and artificial intelligence. In particular, if the subject of the game-theoretic analysis is the Nash equilibrium, then the differentiation can be carried out based on the types of equilibria being played. The latter is obviously possible if there are several equilibria in the game. In particular, the fundamental conceptual idea is the use of a game of the "family dispute" class. Recall that in it, players make a choice between two options. One option is more preferable for player I, the other - for player II. At the same time, if there is an inconsistent choice, then the players receive zero utility. A classic example of this game is given in **Table 1**. The values in the lower left corners of the cells correspond to the utilities of player I, in the upper right - to the utilities of player II.

In the classical formulation of games of this type, the utility values a_{ij} and b_{ij} are set in such a way that the agreed choice **(U, L)** is more favorable for player I, and the agreed choice **(D, R)** is more favorable for player II, i.e.

$$a_{11} > a_{22},\ b_{11} < b_{22} \tag{1}$$

Situations **(D, L)** and **(U, R)** have zero utilities and are disadvantageous to both players. This game has three Nash equilibria. Two in pure strategies: **(U, L)**, **(D, R)**, and one in mixed strategies:

$$((\frac{b_{22}}{b_{11}+b_{22}}, \frac{b_{11}}{b_{11}+b_{22}}),\ (\frac{a_{22}}{a_{11}+a_{22}}, \frac{a_{11}}{a_{11}+a_{22}})) \tag{2}$$

In a mixed equilibrium, the probabilities of choosing pure strategies are determined by the utility ratios. The so-called antagonism of behavior is of fundamental importance. According to it, the mixed strategies equilibrium of each player is determined by the utilities of the opponent.

In the traditional setting of the game called "family dispute", two players are involved, identified as "He" and "She".

This statement of the problem is purely conditional. It can be interpreted more as an ironic parody than as description of some pressing problem of interpersonal relationships. The participants independently choose a meeting place from two options: "football" (**U** or **L** in our notation) or "ballet" (**D** or **R**). The utilities are defined as

$$a_{11} = b_{22} = 3,\ b_{11} = a_{22} = 1. \tag{3}$$

In this case, the mixed equilibrium has the form

$$\{(\frac{2}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3})\}. \tag{4}$$

That is, each player (in the limit) in two-thirds of the cases chooses the option that he likes best, and in one-third of the cases, the option that his opponent likes best. Thus, the mixed equilibrium is symmetrical in terms of the payments received and, accordingly, appears more harmonious. However, the price for symmetry and harmony is that the expected utility of the participants in it is $\frac{2}{3}$, i.e. less than the utility of the losing party in the pure equilibria **(U, L)** and **(D, R)**.

Taking into account the properties given, this game model seems to be a good constructive basis for constructing algorithms for differentiating intelligence types. Effective differentiation in terms of content implies the possibility of associating combinations of players **(NI, NI)**, **(NI, AI)**, **(AI, NI)**, **(AI, AI)** with certain equilibria. At the empirical level, this can be achieved by selecting the corresponding utility values. For example, they should be such that equilibrium **(U, L)** corresponds to the combination of player types **(NI, AI)**, equilibrium **(D, L)** – to the combination **(AI, NI)**, mixed equilibrium – **(NI, NI)**, **(AI, AI)**.

Achieving such a goal requires organizing a series of games that allow statistically significant sequences of strategy choices to be formed by players of different types.

An effective tool for analyzing the behavior of separation game participants can be statistical monitoring of the strategies chosen. It involves comparing the empirical distribution of the strategies actually chosen in the sequences of games (the so-called "empirical mixed strategies") with the theoretical distributions corresponding to a particular mixed strategy, based on statistical criteria of agreement. In the original version, based on the criterion of agreement χ^2. In this case, the conclusion about the degree of closeness of the actual probability distribution to the theoretical one is made based on the *p*-value.

An illustration of the application of this method is presented in **Figure 1**. Within its framework, a study is conducted of the proximity of actually applied mixed strategies to the equilibrium mixed strategy. A pool of players is considered. For each player, the *p*-value is calculated from the sample of strategies used by him, corresponding to the hypothesis that this sample is obtained from the general population with parameters determined by the theoretical equilibrium mixed strategy. The diagram presented in **Figure 1** reflects the distribution of *p*-values obtained in the framework of the experiment. In this case, the proximity of the actual decisions of the experiment participants to the mixed strategy equilibrium of the first player in the game, specified in **Table 1**, was evaluated. For clarity, a trend line is superimposed on the histogram. As you can see, the distribution is clearly shifted to the left, that is, the proportion of "high" *p*-values is significant, at which the hypothesis of the sample belonging to the general population is not rejected. The positive property of the proposed methodology is that it allows us to form a well-founded idea of the "real logic" of

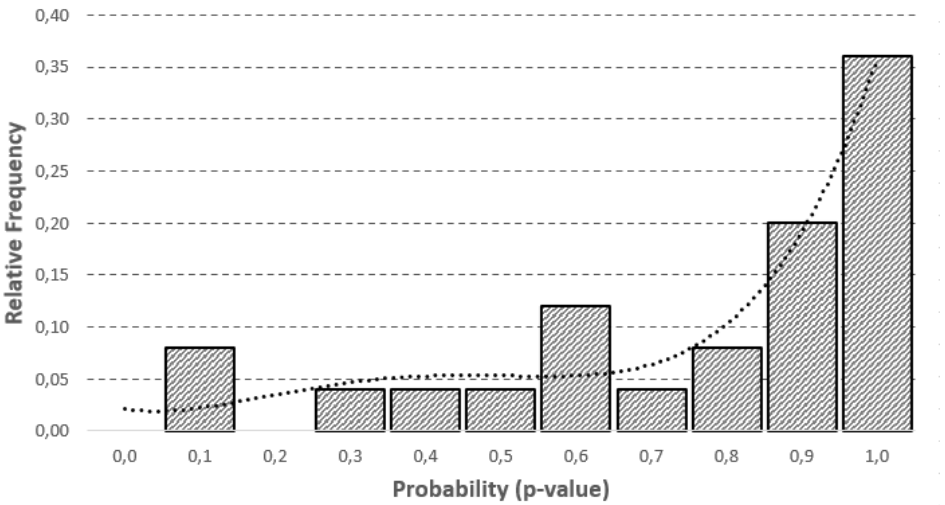

Figure 1. Analysis of p-value for χ^2 criterion, testing the hypothesis about the distribution law of empirical mixed strategies.

the behavior of players (tested intelligent systems), which, generally speaking, may not be consistent with theoretical concepts (Nash equilibrium, etc.).

An important property of the class of games "family dispute" is the consistency of choice and the ability to adapt to a "stubborn opponent". In other words, if player I consistently chooses strategy U when repeating the game, then player II should play strategy **L** (and not **R**). Conversely, if player II persistently plays **R**, then player I would prefer to choose **D**.

As a result, an additional separation feature arises, based on the speed of 'adjustment to the opponent' for artificial and anthropogenic intelligence.

4 Application of Bayesian Games in Solving Differentiation Problems

A promising direction for improving the proposed approach is the transition from games with complete information to games with incomplete information (Bayesian games).

Table 2 presents a diagram of a possible Bayesian modification of the separation game. It is assumed that player I is always a carrier of anthropogenic intelligence (has a single-valued, deterministic type). Player II can be a carrier of both natural (**NI**) and artificial intelligence (**AI**). The type of the second player is unknown to the first. In accordance with the classical formulation of Bayesian games, he has only an idea of the opponent's type: with probability θ, he considers him to be anthropogenic intelligence, with the complementary probability $(1-\theta)$ – artificial.

As is known, strategies in Bayesian games are functions that match the types of players with the actions available to them. In this example, the strategy of player I will be the action he chooses (**U** or **D**). Moreover, this choice will occur taking into account his ideas about the type of player II $(\theta, 1-\theta)$.

The strategy of player II is determined by the pair:

- $\mathcal{A}^{NI}$ – action (**L** or **R**) chosen if player II is natural intelligence;
- $\mathcal{A}^{AI}$ – action (**L** or **R**) chosen if player II is artificial intelligence.

The fundamental subject of study in Bayesian games is the Bayes-Nash equilibrium. In the context of this article, there is no need to give a strict definition of it. If necessary, it can be found in the fundamental work on game theory, [Gibbons, 1992]. We will only emphasize that it defines a stable situation in the game in which it is inappropriate for players (given their existing system of representations) to change the actions that form their strategies, provided that the strategies of their opponents remain unchanged.

Based on the Bayesian game, the following differentiation procedure can be formulated:

- series of games, specified in **Table 2**, is organized;
- empirical statistics of players strategic choices are forming;
- the parameters of the empirical statistical distribution are compared with the theoretical Bayes-Nash equilibrium in the game given by **Table 2**.

In addition to the differentiation capabilities, the game model under consideration is useful from the point of view of describing the rational behavior of carriers of natural intelligence. In particular, on its basis, procedures for correcting probabilistic ideas about the distribution of anthropogenic and artificial intelligence within the framework of the social environment under study can be constructed.

One of the most relevant areas of application of games with incomplete information is the design of economic mechanisms. Its task is to construct such schemes (models, rules) of functioning of socio-economic systems, in which the strategic choice of independent agents acting on the basis of individual utility functions leads to the best result from the point of view of some global function of social choice (common good).

The fundamental ideas of the theory of economic mechanism design were formulated in the works of [Hurwicz, 1973]. They were later developed in the works Myerson [1982], Myerson [1985], Myerson [1983], Maskin and Riley [1984].

Table 2. Bayesian game.

		Player II (NI) L		Player II (NI) R		Player II (AI) L		Player II (AI) R	
Player I	**U**		b_{11}^N		b_{12}^N		b_{11}^A		b_{12}^A
		a_{11}^N		a_{12}^N		a_{11}^A		a_{12}^A	
	D		b_{21}^N		b_{22}^N		b_{21}^A		b_{22}^A
		a_{21}^N		a_{22}^N		a_{21}^A		a_{22}^A	

As is easy to notice, there is a deep semantic relationship between the tasks of mechanism design and the task of differentiation of intelligent systems. The basic scheme of the differentiation procedure from the standpoint of the logic of mechanism design is presented in **Figure 2**. The role of players (agents) is played by intelligent systems, which can be carriers of both natural and artificial intelligence (types NI, AI).

There is some «Referee» to whom the agents communicate their type. The arbitrator in turn must make some decision (outcome) that affects the welfare (utility) of the agents.

If the Referee is a mechanism, this means that there is some function (rule) $\gamma(s_1, \ldots, s_m)$ that uniquely determines the outcome based on the agents' messages. The mechanism is called direct if the outcome function $\gamma(s_1, \ldots, s_m)$ is identical to the social choice function $\Gamma(\circ)$.

In the case of the differentiation problem, the social choice is for all agents (carriers of intelligence) to truthfully inform the arbitrator of their type. It is assumed that their individual utility functions $v_1(\circ), v_2(\circ), \ldots, v_m(\circ)$ are such that in certain situations they can give a better result in the case of deception.

Thus, the task of designing a differentiation mechanism is to find an outcome function $\gamma(\circ)$ such that the choice of agents turns out to be an equilibrium in dominant strategies and corresponds to the global social choice function

$$\gamma(s_1(v_1(\circ)), s_2(v_2(\circ)), \ldots, s_m(v_m(\circ))) = \Gamma(v_1(\circ), v_2(\circ), \ldots, v_m(\circ)) \quad (5)$$

In other words, it is necessary to construct a mechanism in which agents, guided by their individual (selfish) utilities, would report the truth about the type of their intelligence.

The attractiveness of this approach is primarily determined by the fact that it opens up wide possibilities for using a well-developed theory of mechanism design.

At the same time, it allows us to recognize the fundamental theoretical complexities objectively inherent in the problem under study. In particular, the implementation of the social choice function based on the Nash equilibrium (reporting truthful information is an equilibrium) may yield unsatisfactory results. In particular, the equilibrium outcome may be unattractive for the participants in the differentiation procedure. At the same time, the outcome may become attractive if the utility functions are publicly available information, but the rules of the mechanism are unknown.

5 Differentiation Procedures Based on Signal Games

The development of a methodological approach based on Bayesian games may be associated with the transition from static to dynamic game models. In particular, models based on a fairly simple but well-studied class of dynamic Bayesian games (dynamic games with incomplete information) – signal games – appear quite promising and attractive.

Figure 3 shows a basic diagram of the construction of a differentiation procedure based on a signal game. There are two players in the game: the first player is the Sender (sends the signal) and the second player is the Receiver (receives the signal). The Sender can be of two types. In the case of the differentiation procedure, it can be a carrier of natural or artificial intelligence (**NI** or **AI**). In traditional game theory terminology, Nature $\mathcal{N}$ makes an unobservable move, creating the first player of one type or another. This information is unknown to the second player. It is assumed that he can only have an idea of the Sender type, which is given in the form of a probability distribution.

In the scheme shown in **Figure 3**, Sender can send two types of signals (**L** or **R**). In general, there can be more signals.

A natural form of signals in differentiation procedures are messages about the type of intelligence. For example, **L** corresponds to **NI**, **R** – **AI**. Thus, each of the Sender player types can report both true and false information about itself.

The Receiver player reacts to the received signal by making a move **Y** or **N** (believe, don't believe). Based on the results, the players receive payments: $(a_{LY}^N,\ a_{LY}^N)$, $(a_{LN}^N,\ a_{LN}^N)$, etc.

The subject of research in signaling games are the so-called signaling equilibria. A distinction is made between combining equilibria (all types of Senders send the same signal) and separating equilibria (different types of Senders send different signals). It is obvious that, from the point of view of the goals of differentiation procedures, models of signaling games in which separating equilibria exist are of interest.

The fundamental theoretical point for this class of models is that they connect the Nash equilibrium situation (i.e. the situation of strategic choice from which it is inappropriate for players to deviate) with the system of their ideas about each other's types. In the case of differentiation procedures, the ideas must correspond to the objective information-digital level of the analyzed society, i.e. the actual distribution of AI technologies in it.

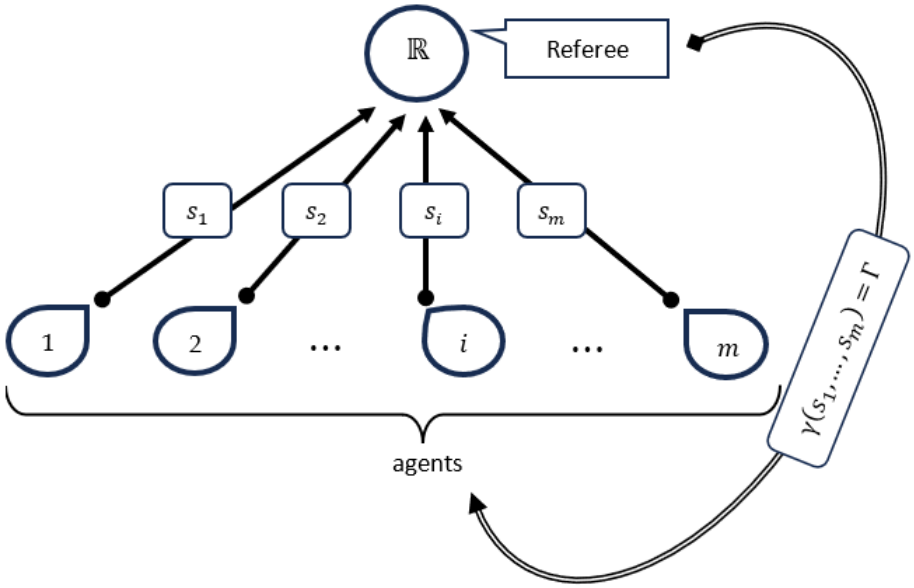

Figure 2. Schematic diagram of the differentiation procedure as a problem of mechanism design.

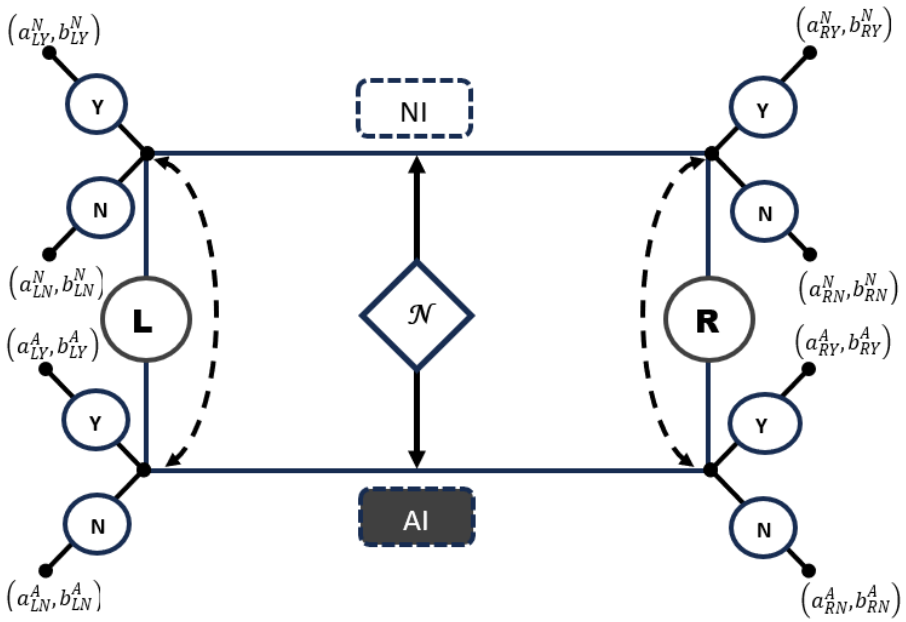

Figure 3. Principal scheme of the signal game.

6 Possible Approaches to Comparative Analysis

When using different differentiation procedures, the problem of their evaluation and comparison inevitably arises. Let us briefly describe one of the possible approaches to its solution.

To obtain reliable and stable results, it is necessary to organize a statistically representative algorithmic process that involves conducting a long series of game differentiation experiments ($k \in \{1, \ldots, K\}$).

Based on the results of each experiment (stage), conclusions are made about the types of players (NI or AI).

We will consider player i to be β_i–separated if the proportion of his correct identifications in the series of experiments conducted is equal to $\beta_i \in [0, 1]$. The overall efficiency of the differentiation procedure ρ can be defined as

$$\alpha_\rho = \min_{i \in \{1,\ldots,m\}} \beta_i^\rho, \tag{6}$$

where β_i^ρ is the proportion of correct differentiation of player i when using the differentiation procedure ρ.

The values of α_ρ become an obvious criterion for assessing the quality (efficiency) of differentiation procedures. In this case, potentially promising areas for studying their properties can be formulated:

- identifying the dependencies of α_ρ on the number of players m;
- identifying dependencies of α_ρ on the number of player strategies in basic games;
- identifying the dependencies of α_ρ on the number of experiments.

7 Discussion and Results

The main result of the conducted research was the system of arguments in favor of the effectiveness and constructiveness of game-theoretic approaches as a tool for solving the problems of separating intelligence types. The principal advantage of this class of methods is that they are free from the weaknesses inherent in expert approaches. Let us recall that the Turing test in its original formulation is a typical example of an expert approach.

The inherent disadvantages of expert methods of differentiation are determined by the fact that experts inevitably extend their system of goals and perceptions to the evaluated object. Thus, a systematic error of imputing to artificial intelligence the properties objectively inherent in human consciousness arises.

When predicting the specificity of artificial intelligence, we should not ignore the fact that it does not necessarily have to be a replica of human intelligence. An analogy with aircraft is admissible here. Indeed, modern airplanes are not clones and replicas of birds. Just as putting wings on a man does not give him the ability to fly.

When studying the problems of intelligence differentiation, it is impossible to ignore the importance of statistical and econometric methods. In particular, the role and possibilities of methods that use the apparatus of cluster analysis. At the same time, it is necessary to recognize that their practical application presupposes some a priori knowledge in which cases the decision was made by anthropogenic and in which cases by artificial intelligence.

Let us emphasize once again that the fundamental premise of this paper is the thesis that the necessary property of artificial intelligence is subjectivity, which is expressed in the presence of an individual system of goals and interests. From this point of view, game-theoretic models act as a tool adequate to the differentiation problem.

Let us remind that game theory is defined as a mathematical discipline dealing with the problems of decision-making under the conditions of conflict of interests. Accordingly, the application of game-theoretic methods and approaches looks promising in terms of the possibilities of separating anthropogenic and artificial intelligence by analyzing conflict situations in which their interests do not coincide. Accordingly, expectations regarding the differences in their behavior and strategic choices are reasonable.

Fundamentally important from the point of view of the tasks of the present study is the thesis about the fundamental substantive closeness of the problems of differentiation of intelligence types to the problems that are the subject of study in the theory of design of economic mechanisms.

One can see fruitful prospects of synergy of theoretical developments obtained in mechanism design in the part of anthropogenic economic agents for the cases of “subjects of a new nature”, i.e. carriers of artificial intelligence. Among other things, these prospects are connected with the tasks of revealing the specificity of interests of “artificial intelligence”, revealing the moments of qualitative transition of “artificial pseudo-intelligence” to “full-fledged artificial intelligence”. At the same time, the tasks of analyzing the logic of “hosts” (owners, manipulators) of AI systems are extremely important and relevant at present.

From the point of view of studying the processes of competition between anthropogenic and artificial intelligence in the medium and long term, fundamental and conceptual forecasts are of undoubted interest. The potential pessimistic scenario does not exclude the situation of “objective loss” of anthropogenic intelligence. This, in turn, gives rise to the problem of prospects for the development of carriers of “natural” (anthropogenic) intelligence and their new place in the society of the future. In particular, we cannot exclude scenarios of shifting the system of value goal-setting towards human physical capabilities, i.e. shifting to those spheres where artificial intelligence is not a rival. Conditionally, this scenario can be called “new kalokagathia”. Let us recall that in ancient Greece the term “kalokagathia” (καλοκαγαθία) meant the ideal of a person who harmoniously combined external (physical) and internal (moral, mental, intellectual) virtues During the long period of development of human civilization, internal (intellectual) virtues invariably had priority in public consciousness. The scenario of a new kalokagathia as a consequence of the civilization consequences of the AI factor, in essence, implies an inversion of the priority. Figuratively speaking, the ironic figurative meaning is replaced by the direct one in the well-known folk wisdom “you have the power - you don’t need the mind”.

It should be emphasized that at the moment such forecasts

are only speculative in nature and cannot claim strict scientific validity. However, certain social phenomena speak in their favor. In particular, we can pay attention to the popularity among modern youth of tattoos, piercings, provocatively long nails, imitating animal behavior. In other words, actions and attitudes oriented to work with the physical characteristics of the body rather than its intellectual improvement are becoming popular. Confirmation or refutation of such hypotheses undoubtedly requires serious and extensive scientific research. Including the methods discussed in this paper.

When speaking about specific forms of implementation of the game-theoretical methods proposed here, we can focus on the following areas. It seems appropriate to conduct a series of tests in student groups. The Herzen State Pedagogical University (Russia, SPb) is considered as a reference university at the preliminary stage. Blind experiments, in which each of the subjects does not know for sure whether he is playing against another student or a neural network, are capable of forming a fairly representative base for verifying the procedures proposed in this paper. In recent years, the branch of psychology known as cyberpsychology has been actively developing. It studies the problems and patterns of interaction between people and the virtual information and software environment. A separate subject area of cyberpsychology is the relationship between people mediated by the virtual information environment. It is quite expected that in the nearest future the development of cyberpsychological directions related to the interaction of natural (anthropogenic) and artificial intelligence, the tasks of identifying (separating) types of intelligence, will occur. Accordingly, game-theoretic methods of separating intelligent systems can find active and direct application within the framework of cyberpsychological tests and procedures.

8 Conclusion

The differentiation algorithms formulated in this paper at the conceptual level undoubtedly need further development and improvement.

Improvement in theoretical terms implies clarification and detailing of algorithms based on static and dynamic Bayesian games. A separate topical set of problems is related to the development of specific differentiation mechanisms, i.e., the development of such "intelligence games" in which a truthful message of its type is an equilibrium in the dominant strategies.

The improvement of differentiation algorithms in applied terms involves the organization of statistically representative series of experiments followed by statistical analysis of the logic and regularities of behavior of carriers of different types of intelligence.

In addition, it is necessary to pay attention to the alternative direction to strategic game models, based on cooperative game-theoretic models. The expectations regarding the constructiveness and fruitfulness of these methods in the research of the regularities of coalition formation, the logic of competition and cooperation between the carriers of anthropogenic and natural intelligence look quite rational.

Declarations

Acknowledgements

Artificial intelligence and neural network tools were not used in the creation of this text.

Competing interests

The author of this paper confirms that he has no competing interests that would preclude its publication.

References

Ali, F., Nakao, Z., and Chen, Y. (2000). Playing the rock paper scissors game with a genetic algorithm. *Proc. 2000 Congr. Evol. Comput. CEC00 (Cat. No. 00TH8512). IEEE*, 1:741–745.

Cenggoro, T., Kridalaksana, A., Arriyanti, E., and Ukkas, M. (2014). Recognition of a human behavior pattern in paper rock scissor game using backpropagation artificial neural network method. *2nd Int. Conf. Inf. Commun. Technol. IEEE*, pages 238–243.

CNews (2023). Chat gpt failed the turing test. `https://www.cnews.ru/news/top/2023-07-28_chatgpt_slomal_test_tyuringa`, Accessed: 16 May 2025.

Dixit, A. and Skeath, A. (1999). *Games of Strategy*. New York: W.W. Norton.

Gang, T., Cho, Y., and Choi, Y. (2017). Classification of rock paper scissors using electromyography and multilayer perceptron. *14th Int. Conf. Ubiquitous Robot. Ambient Intell.*, pages 406–407.

Gardner, H. (2011). *Multiple intelligences: The first thirty years. Introduction to Frames of mind: The theory of multiple intelligences (3rd ed.)*. Basic Books, New York.

Garkusha, N. and Gorodova, J. (2023). Pedagogical opportunities of chatgpt for developing cognitive activity of students. *Vocational Education and Labour Market*, 11(1):6–23. DOI: https://doi.org./10.52944/PORT.2023.52.1.001.

Ghasemi, M., Roshani, G., and Roshani, A. (2020). Detecting human behavioral pattern in rock, paper, scissors game using artificial intelligence. *Computational Engineering and Physical Modeling*, 3(1):25–35.

Gibbons, R. (1992). *Game Theory for Applied Economists*. Princeton University Press.

Hasuda, Y., Ishibashi, S., Kozuka, H., Okano, H., and Ishikawa, J. (2007). A robot designed to play the game "rock, paper, scissors". *Proc. 2000 Congr. Evol. Comput. CEC00 (Cat. No. 00TH8512). IEEE*, pages 2065–2070.

Hu, W., Zhang, G., Tian, H., and Wang, Z. (2019). Chaotic dynamics in asymmetric rock-paper-scissors games. *IEEE Access*, pages 175614–175621.

Hurwicz, L. (1973). The design of mechanisms for resource allocation. *American Economic Review*, 63:1–30.

Ivakhnenko, E. and Nikolskiy, V. (2023). Chatgpt in higher education and science: a threat or a valuable resource? *Vysshee obrazovanie v Rossii*

= *Higher Education in Russia*, 32(4):9–22. DOI: https://doi.org./10.52944/PORT.2023.52.1.001.

Kolyshkin, A., Konyukhovskiy, P., and Yakovleva, T. (2023). Blended educational technologies in the new normalcy. *Obrazovatel'naya Politika = Educational Policy*, 4(96):75–88. DOI: https://doi.org./10.22394/2078–838X-2023–4-75-88.

Konyukhovskiy, P. (2022). Modern models and methods of assessment of the quality of the educational process. *Innovacii = Innovations*, (5(283)):48–58. DOI: https://doi.org./10.26310/2071-3010.2022.284.5.006.

Konyukhovskiy, P. and Kholodkova, V. (2015). Application of game theory in the analysis of economic and political interaction at the international level. *Finansy I Biznes = Finance and Business*, (4):40–57.

Konyukhovskiy, P. and Malova, A. (2013). Game-theoretic models of collaboration among economic agents. *Contributions to Game Theory and Management*, 6:211–221.

Konyukhovskiy, P. and Malova, A. (2015). Stochastic cooperative games application to the analysis of economic agent's interaction. *Contributions to Game Theory and Management*, 8:137–148.

Lianmin, Z., Ying, S., Wei-Lin, C., Hao, Z., Joseph, E., G., and Ion, S. (2023). Chatbot arena: Benchmarking llms in the wild with elo ratings. `https://lmsys.org/blog/2023-05-03-arena/`, Accessed: 16 May 2025.

Maskin, E. and Riley, J. (1984). Optimal auctions with risk averse buyers. *Econometrica*, 52:1473–1518.

Matsumoto, Y., Yamamoto, T., Honda, K., Notsu, A., and Ichihashi, H. (2012). Application of cluster validity criteria to rock-paper-scissors game judgment. *IEEE Int. Conf. Fuzzy Syst.*, pages 1–5.

Mozikov, M., Severin, N., Bodishtianu, V., and Glushanina, M. (2024). Emotional decision-making of llms in strategic games and ethical dilemmas. *38th Conference on Neural Information Processing Systems (NeurIPS 2024)*.

Myerson, R. (1982). Optimal coordination mechanisms in generalized principal-agent problems. *Journal of Mathematical Economics*, 10:67–81.

Myerson, R. (1983). Utilitarianism= egalitarianism and the timing in social choice problem. *Econometrica*, 19(2):883–897.

Myerson, R. (1985). *Bayesian Equilibrium and Incentive Compatibility: an Introduction*. Cambridge University Press.

Salvetti, F., Patelli, P., and Nicolo, S. (2007). Chaotic time series prediction for the game, rock-paper-scissors. *Appl Soft Comput*, 7:1188–96.

Turing, A. (1950). Computing machinery and intelligences. *Mind*, LIX(236):433–460.

Yakovleva, A., Konyukhovskiy, P., and Bi, Y. (2025). The factor of artificial intelligence in hybrid wars. *Journal of Economy and entrepreneurship*, 19(1):453–472.