

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Математические аспекты доверенного искусственного интеллекта

Андрей Нечесов

Центр Искусственного Интеллекта НГУ

SIIS, 2024

NSU and other AI-teams

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



NSU and Institute of Mathematics

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Novosibirsk University & Sobolev Institute of Mathematics



Academician Ershov

AGI



Academician Goncharov



Dr Sviridenko



Dr Vityaev



Dr Nechesov
WeChat: Nechesoff



Dr Pavlovsky



Dr Palchunov



Dr Mancivoda



M.Sc Gumirov

Semantic Programming Strong AI Task approach
Chatbots Trustworthy AI Explainable AI

Digital Twins Neural Networks Generative AI

Smart cities

Smart contracts

Blockchains



Dr Kolonin

NSU: Applied Digital Technology Lab

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Applied Digital Technologies Lab @ Department of Mechanics and Mathematics of NSU



Ivan Bondarenko
researcher



Rustam Mullyadzhonov
PhD, DSc, the head of the lab



Iurii Katser
researcher



Dmitry Morozov
researcher



Chermen Tsgoev
researcher



Ivan Plokhikh
researcher



Rostislav Epifanov
researcher



Yana Fedotova
researcher

- ★ **more than 40 researchers:** mathematicians, physicists, programmers
- ★ **more than 100 peer-reviewed publications** in scientific journals and conference proceedings
- ★ **key expertise:** ML/DL, surrogate modeling, CV in medicine, predictive analytics, NLP, ASR, conversational AI, high-performance computing

Institute of Computation Mathematics and Geophysics

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



Dr O. Klimenko



D. Kosyakov



Dr A. Guskov



Prof M. Marchenko



Prof V. Barakhnin



R. Permyakov



Dr A. Ivanov

- Machine learning
- Software solutions
- Cybersecurity
in applications
- Seismics
- Chemistry
- Bioinformatics
- Scientometrics
and others



I. Marinin



M. Gorodnichev



S. Kireev



Dr A. Yakimenko



Dr Yu. Medvedev



Dr V. Perepelkin



N. Podkolodny



Laboratory of Artificial Intelligence and Information Technologies
@ Institute of Computational Mathematics and Mathematical Geophysics SB RAS

Institute of Informatics Systems

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

AI Directions@Institute of Informatics Systems

Russian Academy of Sciences

NLP
Ontology Engineering
Knowledge Graphs



Prof. Zagorulko Yuri



Prof. Apanovich Zina

AI for DB
Reinforcement Learning
Rule Based ML



Dr. Ponomaryov Denis

Software Verification
Trustworthy AI



Dr. Anureev Sergey



Dr. Seryl Alexei



Dr. Sidorova Elena



Dr. Batura Tanya



Dr. Demin Alexander



Zinchenko Sergey



Dr. Garanina Natasha



Dr. Kondratyev Dmitry



Dr. Maryasov Ilya



Dr. Bruches Elena



Shestakov Vladimir

If you need

- Safe software
- Good quality of requirements

Contact us!

Ready to discuss cooperation

Verification Work Group, IIS SB RAS



PhD Anureev



PhD Garanina



PhD Kondratyev

Methods

- Deductive Verification
- Model Checking
- Runtime Verification
- Finite State Machines
- Program Logics
- Ontologies and Logics of Knowledge
- AI-based Reasoning

Application scopes

- Hardware Verification
- Software Verification
- AI Verification
- Verification of AI-based Autonomous Systems
- Embedded Software Verification
- Control Software Verification
- Development of Programming Languages and Tools for Control Software
- Development of Software Requirements
- Development of Formal Models for Software and Hardware

Contacts

- Igor Anureev, Head of the Group, +79137694105, anureev@gmail.com
- Natalia Garanina, +79139189293, natta.garanina@gmail.com
- Dmitry Kondratyev, +79232484459, apple-66@mail.ru

Call, write, we find time to meet

Institute of Informatics Systems

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Ugra Research Institute of Information Technologies

Generative AI



Valerian Abbasov

Digital avatar



Oleg Sokolov

Neural Networks

Causal language
modeling

Large Language
Model (LLM)



Mikhail Rusanov



Ivan Yaryshkin

Summarization

+7 (3467) 360-100

Strong AI



Andrey Melnikov
D.Eng., Prof.



Yuri Polishchuk
D.Sc., Prof.

OFFICE@URIIT.RU

Prompt
engineering



Alexander
Tsaregorodtsev, PhD



Gleb Kochergin, PhD

URIIT.RU

Retrieval-Augmented
Generation (RAG)

Question
answering



Odiljon Negmatulloev



Vladimir Krupitsky

LLM chain

Natural language
processing (NLP)

Chatbot

URIIT.RU

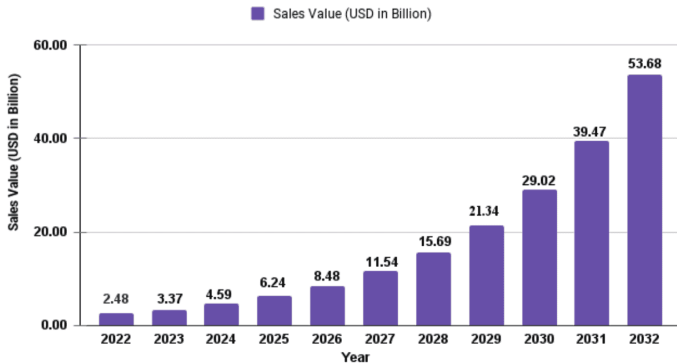
AI Education World Market

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



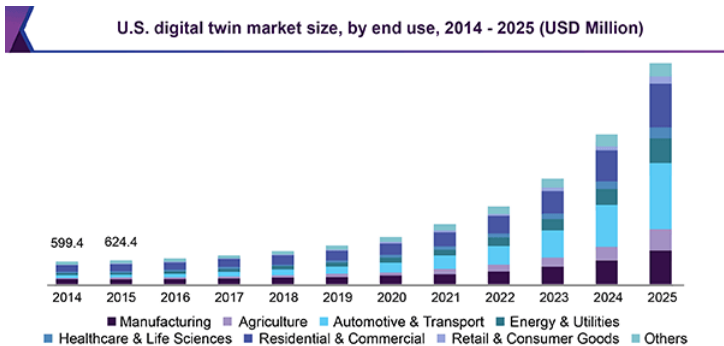
US Digital Twin Market

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



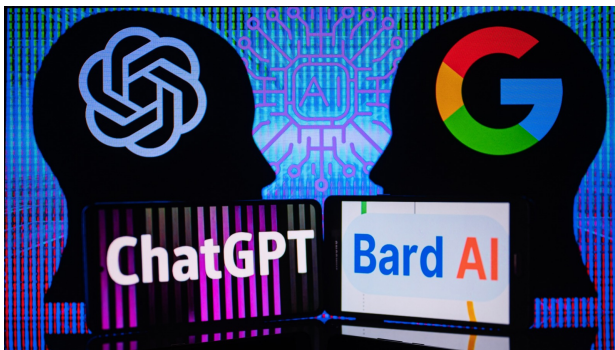
Source: www.grandviewresearch.com

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



Data

Genomics



Proteomics



Metabolomics



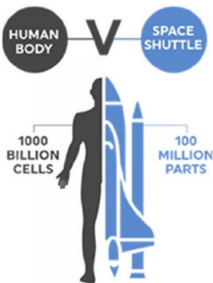
Clinical data



Clinical history



Life style

*Computational models
of Digital Twins**Results enabled by
Digital Twins*Better
TherapiesPromoting
WellbeingPERSONAL
PROFILE

Saving cost

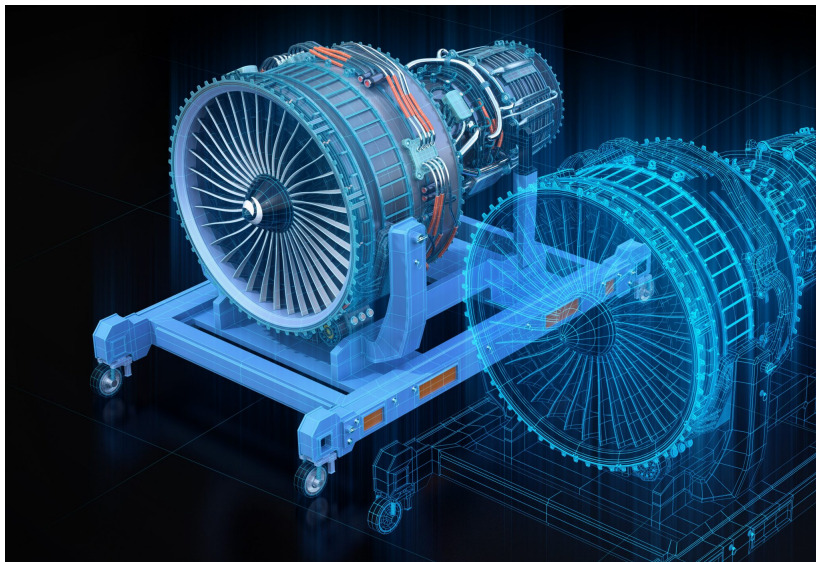
Finer
DiagnosticsEnabling
Prevention

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

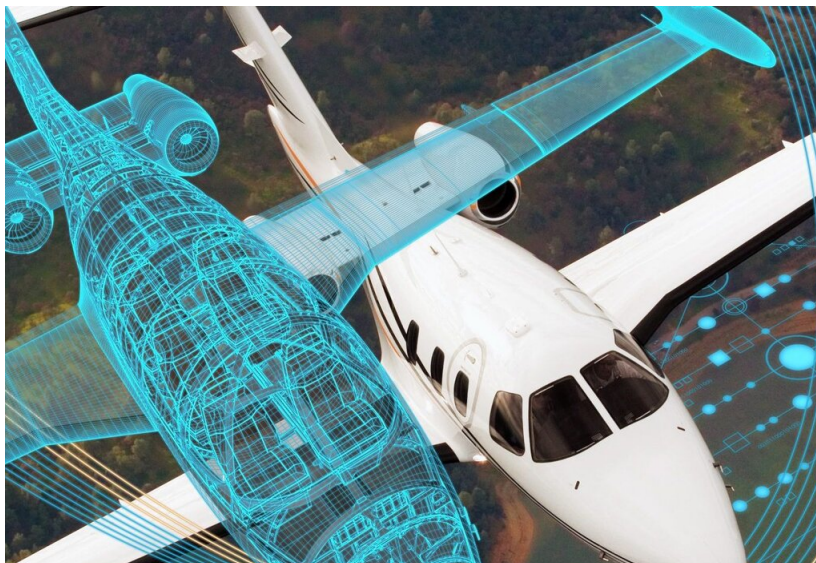


AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



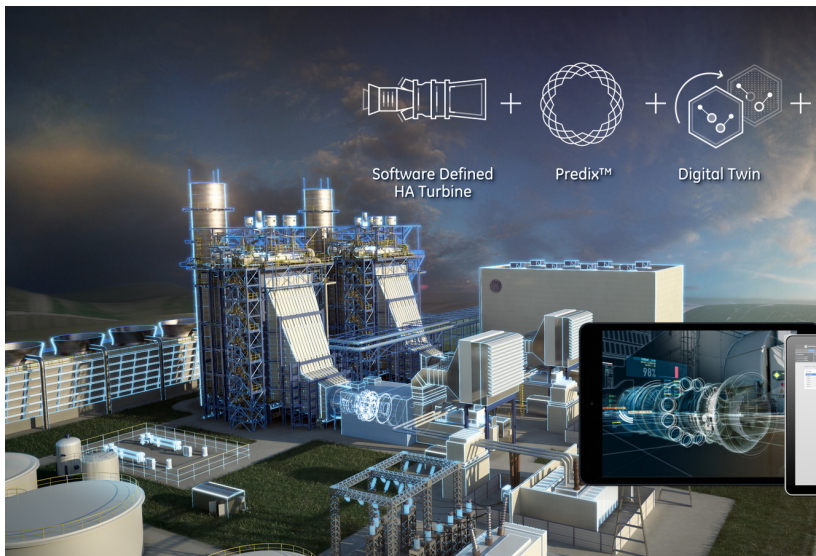
Digital factory

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

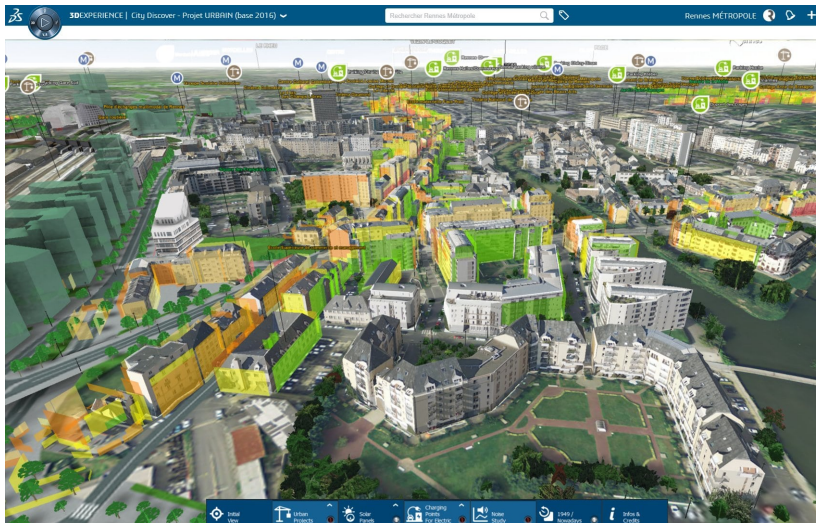


AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



Алиса в стране чудес

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

«Задача здравого смысла», названная в статье «проблемой AIW», на самом деле проста: «У Алисы есть N братьев и M сестер. Сколько сестер у брата Алисы?» Значения N и M (всегда натуральные числа) и порядок упоминания братьев и сестер могут различаться. В целом, у языковых моделей (LLMs) средняя частота правильных ответов была значительно ниже 50 процентов, при этом более крупные модели, как правило, работали заметно лучше, чем модели меньшего размера (например, GPT-4 показывает частоту правильных ответов чуть выше 60 процентов).

ХАИ - 4 принципа: Сентябрь 2021

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

The National Institute of Standards and Technology (NIST)

- 1 **Explanation:** ИИ должен предоставить объяснение, которое при необходимости можно детализировать
- 2 **Meaningful:** Объяснение ИИ должно быть понятно пользователю
- 3 **Accuracy:** Объяснение должно согласовываться с алгоритмами системы
- 4 **Limits:** ХАИ не должен выходить за рамки своих ограничений

Пекин задал курс на ХАИ + Тех. Незав.

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Cyberspace Administration of China, САС опубликовала 10 мая 2023 правила для ИИ-компаний и их разработок.

Суть правил:

Компании будут нести ответственность за своих чат-ботов, если:

- первоначальные источники данных нелегитимны
- за все результаты выдачи чат-ботов
- за контент, который нарушал бы основные принципы социалистической партии
- за призывы к свержению режима или действия порочащие партию

Проблемы современного ИИ

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

- Проблема черного ящика
 - нет прозрачности
 - нет проверяемости
 - нет объяснимости
- Проблема централизации
- Проблема аудита и верификации
- Проблема надежности
- Проблема верификации
- Проблемы безопасности
- Проблема обучения

ChatGPT + Wolfram Alpha

Alpha Geometry: нейросети + правила Хорна
 $P_1(\bar{x}), \dots, P_n(\bar{x}) \rightarrow Q(\bar{x})$

5 key technologies

- Блокчейны/Мультиблокчейны и другие защищенные БД
- р-вычислимые Смарт Контракты
- Гибридный ИИ: Теория обучения + логический вывод
- Контроллеры
- Внутренняя безопасность + Верификаци

Список работ:

- Nechesov, A.; Safarov, R. WEB 3.0 AND SMART CITIES. 2017.
- Goncharov, S.; Nechesov, A. Axiomatization of Blockchain Theory. Mathematics 2023, 11, 2966
- Goncharov, S.; Nechesov, A. AI-Driven Digital Twins for Smart Cities. Eng. Proc. 2023, 58, 94.
- Nechesov, A.; Novikov, S. Axiomatization of Multiblockchain Theory. (is written)

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

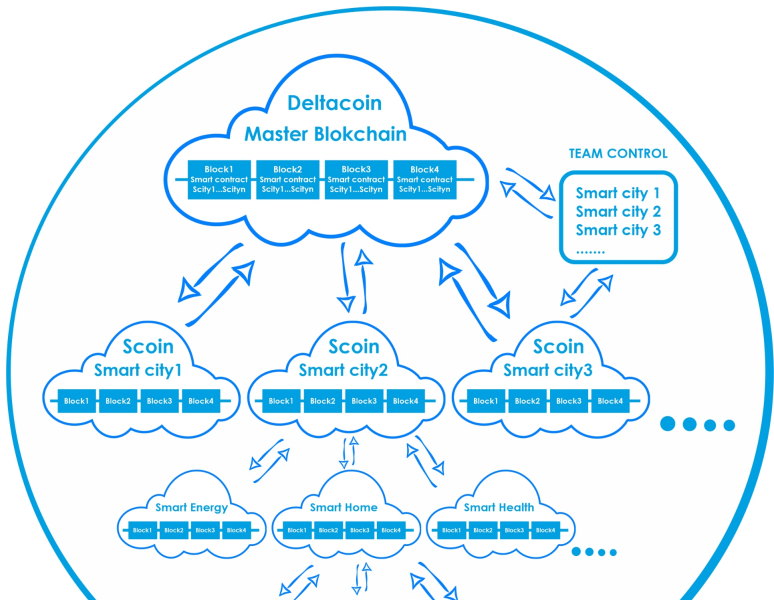


AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ



Список работ:

- Goncharov, S.; Sviridenko D.; Vityaev E. Task Approach to Artificial Intelligence. CEUR Workshop Proceedings. 2020. V.2642. P.1-6.
- Vityaev, E.; Goncharov, S.; Sviridenko, D. About the task approach in artificial intelligence. Siberian Philosophical Journal 2020. v18(2) pp.5-29.
- Vityaev, E.; Goncharov, S.; Gumirov, V.; Mantsivoda, A.; Nechesov, A.; Sviridenko, D. Task approach: on the way to trusting artificial intelligence. WORLD CONGRESS SYSTEMS THEORY, ALGEBRAIC BIOLOGY, ARTIFICIAL INTELLIGENCE: Mathematical Foundations and Applications SELECTED WORKS. 2023

ИСТОРИЯ СТАНОВЛЕНИЯ ЗАДАЧНОГО ПОДХОДА



Акад. СССР **А.Н.Колмогоров** – исчисление задач как интерпретация интуиционистского исчисления высказываний (1932 г.), задачный подход к школьному образованию (50-60-е гг., XX век)



Г.С.Альтшуллер – автор ТРИЗ, задачный подход к инженерии (70-е гг., XX век – н.вр.)



Акад. РАН **Ю.Л.Ершов, д.ф.н. К.Ф.Самохвалов** – задачный подход к основаниям математики (70-80-е гг. XX века)



Акад. РАН **С.С.Гончаров, д.ф.-м.н. Е.Е.Вит'ев, д.ф.-м.н. Д.И.Свириденко** – задачный подход к искусственному интеллекту (80-е гг., XX век – н.вр.)

$(F(x, y), y = t(x), p)$ – это вероятностное знание

где

$F_i(x, y) : \forall x \exists y \Phi_i(x, y) \rightarrow \Psi_i(x, y)$

Иерархия знаний

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Иерархия знаний:

Мы можем говорить об иерархии знаний (\leq_{φ})

$$(F_1(x, y), y = t_1(x), p_1) \leq_{\varphi} (F_2(x, y), y = t_2(x), p_2)$$

\Leftrightarrow

- $\Phi_1 \subseteq \& \Phi_2$
- $\Psi_1 \subseteq \& \Psi_2$
- $p_1 \leq p_2$

Теория обучения и иерархия знаний

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Если мы хотим понять насколько эффективно решение $y = t(x)$, мы подставляем его вместо y :

$$F_i(x, t(x)) : \forall x \Phi_i(x, t(x)) \rightarrow \Psi_i(x, t(x))$$

и проверяем истинность этого утверждения на фактах, которыми мы обладаем

Оценка эффективности решения

AI

Логическая формула (Задача) + Потенциальное решение

$$\forall x \exists y \Phi(x, y) \rightarrow \Psi(x, y), y = t(x)$$

База данных содержит все проблемы с решениями

$$F_k(x, y) : \exists y \Delta_k(c_i, y) \rightarrow \Theta_k(c_i, y), y = t(c_i)$$

где $\Phi \subseteq_{\&} \Delta_k$

Probability value

Мы перебираем факты и считаем вероятность. Если фактов слишком много, то используем один из методов при работе с большими данными, например, RSP.

$$p(\Psi(x, t(x)) \mid \Phi(x, t(x))) = \frac{\sum_{i,k \in K} \mu(\Psi(c_i, t(c_i)))}{\sum_{i,k \in K} \mu(\Delta_k(c_i, t(c_i)))}$$

$\mu(\Phi(c_i, d_j)) = 1$, если $\Phi(c_i, d_j)$ - истина и 0 иначе.

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Теорема (о поиске наилучшего решения)

Пусть B - база вероятностных знаний, F - некоторая задача, которую необходимо решить. Тогда существует алгоритм поиска наилучшего решения сложность которого полиномиальна.

Лемма (о проблеме логического вывода)

Пусть $\forall x A(x, t(x)) \rightarrow B(x, t(x))$ и $\forall x B(x, t(x)) \rightarrow C(x, t(x))$
- два вероятностных знания с эффективностью решений
близких к 1, но можно подобрать такие задачи и их
решения, что $\forall x A(x, t(x)) \rightarrow C(x, t(x))$ будет любым
числом из интервала $(0, 1)$.

Языки программирования для алгоритмов ИИ

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

- почти все языки Тьюринг-полные Python, C++, Java и т.д.
- Проблема остановки для Тьюринг полных языков!
- Опасность зацикливания и зависания вычислимых ИИ-реализаций.
- Проблема гарантированной точности (академик С.К.Годунов, далее академик С.С.Гончаров)

РАГ-теорема

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Вводится понятие GNF-системы системы как n -ки вида $(\Sigma, \sigma, \Omega, L, HW(\mathfrak{M}), P, F, G)$ и определяется r -вычислимая GNF-система.

РАГ-теорема Гончаров, Нечесов 2021

Пусть G – r -вычислимая GNF-система, тогда наименьшая неподвижная точка Γ^* оператора $\Gamma_{P_1, \dots, P_n}^{HW(\mathfrak{M})}$ является r -вычислимой.

Goncharov, Nechesov 2021

Polynomial Analogue of Gandy's Fixed Point Theorem.

<https://doi.org/10.3390/math9172102>

Решение проблемы $P = L$

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Гончаров, Свириденко и Нечесов предложили ряд расширений базого языка.

$$L_0 \subseteq L_1 \subseteq L_2 \subseteq \dots L_n \subseteq L = P$$

Solution of the Problem $P = L$, 2022

Goncharov, Nechesov <https://doi.org/10.3390/math10010113>

Пусть $\mathbb{HW}(\mathfrak{M})$ – p -вычислимая модель сигнатуры σ , тогда:

- 1) сложность любой L -программы является полиномиальной.
- 2) для любой p -вычислимой функции существует подходящая L -программа реализующая ее.

Объектно-ориентированный ρ -полный язык L^*

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

На базе языка L , построен объектно-ориентированный логический язык программирования L^* , а также реализован механизм исполнения L^* -программ с помощью разработанной нами виртуальной машины V .

Теорема (о консервативности расширения)

Язык L^* является консервативным расширением ρ -полного языка L .

Замечание: Язык L^* готов к применению для реализации алгоритмов искусственного интеллекта!

Goncharov, Nechesov 2022

Semantic programming for AI and Robotics

<https://doi.org/10.1109/SIBIRCON56155.2022.10017077>

Нами разработана методология программирования в высокоуровневых языках:

- по Тьюринг-полному языку L строится специальное r -полное обеднение L_M
- запрещаем в циклах FOR переприсваивание инкремента
- убираем другие операторы цикла
- оставляем рекурсивные определения только удовлетворяющие требованиям Ганди

Теорема [Goncharov, Nechesov, Sviridenko, 2024]
(о существовании ρ -полного подязыка)

Пусть L язык удовлетворяет критериям как сильной выразительности, так и ρ -полного обеднения тогда существует алгоритм полиномиальной вычислительной сложности проверяющий любую программу в этом языке на согласованность с методологией программирования.

Теорема [Goncharov, Nechesov, Sviridenko, 2024] (о методологии программирования)

Пусть L язык удовлетворяет критериям как сильной выразительности, так и p -полного обеднения тогда любая программа на этом языке, соответствующая методологии программирования, имеет полиномиальную вычислительную сложность.

Функциональный вариант РАГ-теоремы

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Вводится понятие FGNF-системы системы как n -ки вида $(\Sigma, \sigma, \Omega, L, HW(\mathfrak{M}), F, T, G)$ и определяется r -вычислимая FGNF-система:

- если f_j входит в t_k то только в виде $f_j(x_i)$
- $f_j(x_i)$ и $f_k(x_i)$ не могут входить одновременно ни в какой t_j
- требование единственности порождения $\neg(t_i(w_1, \dots, w_n) \downarrow \& t_j(w_1, \dots, w_n) \downarrow), i \neq j$
- для любого $t_j^{\bar{y}}(\bar{x}, \bar{y})$ не верно что существует $i \in N$ т.ч. $x_i \in V(t_j^{\bar{y}}(\bar{x}, \bar{y}))$ и $y_i \in V(t_j^{\bar{y}}(\bar{x}, \bar{y}))$
- $\forall T_k \exists C, \rho \forall t_j \in T_k t(t_j^{\bar{y}}(\bar{x}, \bar{y})) \leq C \cdot (|\bar{x}| + |\bar{y}|)^\rho$
- если $V_y(t_j^{\bar{y}}(\bar{x}, \bar{y})) \neq \emptyset$ тогда $|t_j^{\bar{y}}(\bar{w}, \bar{l})| \leq |\bar{w}| + |\bar{l}|, t_j \in T_k$

Функциональный вариант РАG-теоремы

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Теорема (FРАG-теорема, 2024)

Пусть задана p -вычислимая FGNS-система с начальными p -вычислимыми функциями f_1, \dots, f_n тогда наименьшая неподвижная точка $\Gamma_{\bar{f}}^*$ оператора $\Gamma_{f_1, \dots, f_n}^{\langle HW(\mathfrak{M}), \sigma \rangle}$ является p -вычислимой.

проекты ЦИИ НГУ

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

- Платформа по быстрому развертыванию ЦД умных городов
- Доверенный модуль обучения систем искусственного интеллекта
- Интеллектуальная система контроля внимания при дистанционном обучении
- Системы мониторинга и диагностики заболеваний
- Безопасность ИТ-инфраструктуры городской среды
- Разработка системы детектирования событий в видеопотоке
- Разработка и эксплуатация edge-устройств в системах «Умного города»

AI

Nechesov

AI and
Digital
Twins

Проблемы
ИИ

Спасибо за внимание!

Андрей Витальевич Нечесов
Telegram, WhatsApp: +7-913-480-37-63
Email: nechesoff@gmail.com