ISSN 0235-1188 (Print)
ISSN 2618-8961 (Online)

64(1)` 2021

ФИЛОСОФСКИЕ НАУКИ

ФН

RUSSIAN JOURNAL OF PHILOSOPHICAL SCIENCES

- *КРЕМНИЕВАЯ ЦИВИЛИЗАЦИЯ*
- *СУРРОГАТНАЯ ЖИЗНЬ*
- *ПОЗНАВАТЕЛЬНЫЙ КРИЗИС*
- *ТВОРЧЕСТВО НЕЙРОСЕТЕЙ*
- *ЦИФРОВОЕ БЕССМЕРТИЕ*
- *ДЕМИУРГ ВСЕЛЕННОЙ*
- *ВРЕМЯ ФИЛОСОФИИ!*

МОСКВА
ГУМАНИТАРИЙ

MOSCOW
HUMANIST PUBLISHING HOUSE

МИНИСТЕРСТВО НАУКИ
И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

АКАДЕМИЯ
ГУМАНИТАРНЫХ
ИССЛЕДОВАНИЙ

ФН

ФИЛОСОФСКИЕ НАУКИ

Том 64. № 1/2021

Научный образовательный просветительский журнал
Издается с 1958 года. Выходит ежемесячно

Журнал включен в «Реферативный журнал» и в базы данных
ВИНИТИ РАН, **Российский индекс научного цитирования,**
EBSCO Academic Complete Search, The Philosopher's Index.
Сведения о журнале ежегодно публикуются в международной
справочной системе по периодическим и продолжающимся изданиям
«**Ulrich's Periodicals Directory**»

DOI: 10.30727/0235-1188

Москва
Гуманитарий

Международная редакционная коллегия

Председатель международной редакционной коллегии **Гусейнов А.А.**

Амплеби С.А., д.филос., проф. Ун-та Джорджа Вашингтона (США); **Бараш Дж.Э.**, д.филос., проф. Ун-та Пикардии им. Ж. Верна (Франция); **Брес И.**, проф., гл. ред. «Revue philosophique de la France et de l'étranger» (Франция); **Даллмайр Ф.Р.**, д.филос., проф. Ун-та Нотр-Дам (США); **Денн М.,** д.филос., проф., координатор отношений и международных проектов с Россией и странами Восточной Европы Ун-та Бордо им. Монтеня (Франция); **Майнцер К.**, проф. Мюнхенского технического университета, научный директор Академии им. Карла фон Линде (Германия); **Тиханов Г.**, д.филос., зав. кафедрой сравнительного литературоведения Лондонского ун-та Королевы Марии (Великобритания); **Чжан Байчунь,** проф. Пекинского педагогического университета (Китай); **Штольценберг Ю.**, д.филос., проф. Ун-та Галле-Виттенберга им. М. Лютера (Германия); **Эпштейн М.**, проф. Ун-та Эмори (США), директор Центра обновления гуманитарных наук Даремского ун-та (Великобритания); **Эспехо Р.**, д.филос., президент Международной организации систем и кибернетики (Великобритания), приглашенный проф. Ун-та Сантьяго (Чили).

Редакционная коллегия

Председатель редакционной коллегии **Смирнов А.В.**

Автономова Н.С., ак. Академии гуманитарных исследований (АГИ), д.филос.н., гл.н.с. ИФ РАН; **Алексеев П.В.**, д.филос.н., проф. МГУ им. М.В. Ломоносова; **Апресян Р.Г**., ак. АГИ, д.филос.н., зав. сектором ИФ РАН; **Аршинов В.И.**, д.филос.н., гл.н.с. ИФ РАН; **Блауберг И.И.**, д.филос.н., в.н.с. ИФ РАН; **Вдовина И.С.**, д.филос.н., гл.н.с. ИФ РАН; **Водолазов Г.Г.**, ак. Академии политической науки (АПН), вице-президент АПН, проф. МГИМО (У); **Губин В.Д.**, ак. АГИ, д.филос.н., зав. кафедрой истории зарубежной философии РГГУ; **Гусейнов А.А.**, ак. РАН, научный руководитель ИФ РАН; **Давыдов А.П.**, д.культурологии, гл.н.с. ИС РАН; **Доброхотов А.Л.**, ак. АГИ, д.филос.н., проф. НИУ ВШЭ; **Дубровский Д.И.**, д.филос.н., проф., гл.н.с. ИФ РАН; **Журавлев А.Л.**, ак. РАН, научный руководитель ИП РАН; **Лепский В.Е.,** д.психолог.н., проф., гл.н.с. ИФ РАН; **Михайлов И.А.**, к.филос.н., с.н.с. ИФ РАН; **Мотрошилова Н.В.**, д.филос.н., гл.н.с. ИФ РАН; **Пантин В.И.**, ак. АПН, зав. отделом ИМЭМО РАН; **Пивоваров Ю.С.**, ак. РАН, научный руководитель ИНИОН РАН; **Порус В.Н.**, д.филос.н., руководитель Школы философии фак-та гум. наук НИУ ВШЭ; **Розин В.М.**, д.филос.н., г.н.с. ИФ РАН; **Рябов В.В.**, чл.-кор. РАО, президент МГПУ; **Северикова Н.М.,** к.филос.н., заслуж.н.с. МГУ им. М.В. Ломоносова; **Сиземская И.Н.**, д.филос.н., гл.н.с. ИФ РАН; **Смирнов А.В.,** ак. РАН, академик-секретарь Отделения общественных наук РАН, директор ИФ РАН; **Смолин О.Н.**, ак. РАО, 1-й зам. пред. комитета ГД ФС РФ по образованию и науке; **Степанянц М.Т.**, ак. АГИ, д.филос.н., зав. кафедрой ЮНЕСКО ИФ РАН; **Тульчинский Г.Л.**, д.филос.н., проф. НИУ ВШЭ (СПб); **Турбовской Я.С.**, д.пед.н., председатель Координационного совета Института стратегии развития образования РАО; **Федотова В.Г.**, ак. РАЕН, д.филос.н., гл.н.с. ИФ РАН; **Черниговская Т.В.,** чл.-корр. РАО, д.б.н., д.ф.н., проф., зав. кафедрой СПбГУ; **Шевченко В.Н.**, д.филос.н., гл.н.с. ИФ РАН.

Редакция:

Главный редактор Дубровский Д.И.

Научный редактор Винник Д.В.

Редактор отдела культурологии и религиоведения Дуркин Р.А.

Литературный редактор Тукузова Т.М.

Верстка: Топилина В.М.

E-mail: academyRH@list.ru

http://www.phisci.info

Шеф-редактор Мариносян Х.Э.

ISSN 0235-1188 (Print)

ISSN 2618-8961 (Online)

© Философские науки, 2021

© Академия гуманитарных исследований, 2021

© Издательский дом «Гуманитарий», 2021

THE MINISTRY OF SCIENCE
AND HIGHER EDUCATION OF
THE RUSSIAN FEDERATION

ACADEMY FOR
RESEARCH INTO
THE HUMANITIES

ΦН

RUSSIAN JOURNAL OF PHILOSOPHICAL SCIENCES (FILOSOFSKIE NAUKI)

Vol. 64. No. 1/2021

Scientific and Educational Journal
Published since the Year 1958. Issued Monthly

The journal is listed in the ***Abstracts Journal*** and **databases of the VINITI** (All-Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences), **Russian Index of Science Citation, EBSCO Academic Complete Search, The Philosopher's Index.**
The information about the journal is published annually in the international information system on serial publications ***Ulrich's Periodicals Directory***

DOI: 10.30727/0235-1188

Moscow
Humanist Publishing House

International Editorial Board:
Chairman of the International Editorial Board **Guseinov A.A.**
Barash J.A., Dr., Prof., Jules Verne University of Picardy (France); **Brès Y.**, Dr., Prof. em., Paris Diderot University – Paris 7, Editor-in-Chief of the *Revue philosophique de la France et de l'étranger* (France); **Dallmayr F.R.**, Ph.D., Packey J. Dee Professor at the University of Notre Dame (USA); **Dennes M.**, Dr., Prof., coordinator of relations and international projects with Russia and other Eastern European countries, Montaigne Bordeaux 3 University (France); **Epstein M.**, Ph.D., S.C. Dobbs Prof. at Emory University (USA), Director of the Centre for Humanities Innovation at Durham University (UK); **Espejo R.**, Ph.D., President of the World Organisation of Systems and Cybernetics (UK), Visiting Prof. at the University of Santiago (Chili); **Mainzer K.**, Dr., Prof. em., Technical University of Munich (Germany); **Stolzenberg J.**, Dr., Prof. em., Martin Luther University of Halle-Wittenberg (Germany); **Tihanov G.**, Ph.D., George Steiner Professor of Comparative Literature at Queen Mary University of London (UK); **Umpleby S.A.**, Ph.D., Prof., The George Washington University (USA); **Zhang Baichun,** Prof., Beijing Normal University (China).

Editorial Board:
Chairman of the Editorial Board **Smirnov A.V.**
Alexeev P.V., D.Sc., Prof., Lomonosov Moscow State University (MSU); **Apressyan R.G.**, D.Sc., Head of the Department, Institute of Philosophy of the Russian Academy of Sciences (IPhRAS); **Arshinov V.I.**, D.Sc., Prin.Res.Fell., IPhRAS; **Avtonomova N.S.**, D.Sc., Prin.Res. Fell., IPhRAS; **Blauberg I.I.**, D.Sc., Lead.Res.Fell., IPhRAS; **Chernigovskaya T.V.,** RAE Corr Memb., D.Sc., Prof., Head of the Department at Saint Petersburg State University (Saint Petersburg); **Davydov A.P.**, D.Sc., Prin.Res.Fell., Institute of Sociology of the RAS; **Dobrohotov A.L.**, D.Sc., Prof., National Research University Higher School of Economics (NRU HSE); **Dubrovsky D.I.,** D.Sc., Prof., Prin.Res.Fell., IPhRAS; **Fedotova V.G.**, RANS Full Memb., Prin.Res.Fell, IPhRAS, Ph.D.; **Gubin V.D.**, D.Sc., Head of the Department of the History of Foreign Philosophy at the Russian State University for the Humanities; **Guseynov A.A.**, RAS Full Memb., Director of the Institute of Philosophy of the RAS; **Lepskiy V.E.**,D.Sc., Prof., Prin.Res.Fell., IPhRAS; **Mikhaylov I.A.**, Ph.D., Sen.Res.Fell., IPhRAS; **Motroshilova N.V.**, D.Sc., Prof., Prin.Res.Fell., IPhRAS; **Pantin V.I.**, Academy of Political Sciences (Russia) Full Memb., Head of the Department at the Institute of World Economy and International Relations of the RAS; **Pivovarov Yu.S.**, RAS Full Memb., Scientific Director of the Institute of Scientific Information for Social Sciences of the RAS; **Porus V.N.**, D.Sc., Head of the Department at the NRU HSE; **Rozin V.M.**, D.Sc., Lead.Res.Fell., IPhRAS; **Ryabov V.V.**, RAE Corr. Memb., President of the Moscow City Teacher Training University; **Severi-kova N.M.**, Ph.D., Honour.Res.Fell., Faculty of Philosophy, MSU; **Shevchenko V.N.**, D.Sc., Prin.Res. Fell., IPhRAS; **Sizemskaya I.N.**, D.Sc., Prin.Res.Fell., IPhRAS; **Smirnov A.V.,** RAS Full Memb., Academician-Secretary of the Department of Social Sciences of the RAS, Director of the Institute of Philosophy of the RAS; **Smolin O.N.**, RAE Corr. Memb., Fisrt Deputy Chairman of the Russian State Duma Committee for Education; **Stepanyants M.T.**, D.Sc., Prin.Res.Fell., UNESCO Chairholder at the IPhRAS; **Tulchinskii G.L.**, D.Sc., Prof., NRU HSE (St. Petersburg); **Turbovskoy Ya.S.**, D.Sc., Chairperson of the Coordination Council at the Institute for the Theory and History of Pedagogy of the RAE; **Vdovina I.S.**, D.Sc., Prin.Res.Fell., IPhRAS; **Vodolazov G.G.**, APS Full Memb., Vice President of the Academy of Political Sciences (APS), Prof., Moscow State Institute of International Relations; **Zhuravlev A.L.**, RAS Full Memb., Chief Scientific Officer of the Institute of Psychology of the RAS.

Editorial Staff:
Main Editor Dubrovsky D.I.
Scientific Editor Vinnik D.V.
Cultural and Religious Studies Department's Editor Durkin R.A.
Literary editor Tukuzova T.M.
Page Layout: Topilina V.M.
E-mail: academyRH@list.ru
http://www.phisci.info
Editor-in-Chief Marinosyan Kh.E.

ISSN 0235-1188 (Print)
ISSN 2618-8961 (Online)

© Russian Journal of Philosophical Sciences, 2021
© Academy for Research into the Humanities, 2021
© Humanist Publishing House, 2021

ОГЛАВЛЕНИЕ

ПЕРСПЕКТИВЫ ЧЕЛОВЕЧЕСТВА. ФИЛОСОФИЯ ГУМАНИТАРНО-ТЕХНОЛОГИЧЕСКОГО РАЗВИТИЯ

Философия искусственного интеллекта ■

В.А. ЛЕКТОРСКИЙ — О философских проблемах искусственного интеллекта и когнитивных исследований 7

Д.И. ДУБРОВСКИЙ — Задача создания Общего искусственного интеллекта и проблема сознания 13

Т.С. АХРОМЕЕВА, Г.Г. МАЛИНЕЦКИЙ — Новое пифагорейство, сложность, искусственный интеллект 45

Е.А. БЕЗЛЕПКИН (Новосибирск), А.С. ЗАЙКОВА (Новосибирск) — Нейрофилософия, философия нейронаук и философия искусственного интеллекта: проблема различения 71

Векторы развития искусственного интеллекта ■

В.Е. ЛЕПСКИЙ — Искусственный интеллект в субъектных парадигмах управления 88

A.V. SHILLER, O.E. PETRUNYA — Architectural Approach to Design of Emotional Intelligent Systems

{А.В. ШИЛЛЕР, О.Э. ПЕТРУНЯ — Архитектурный подход к созданию эмоциональных интеллектуальных систем} 102

КОГНИТИВНОЕ ПРОСТРАНСТВО

Электронная культура: проблемы и перспективы ■

В.Г. БУДАНОВ, А.Р. ЕФИМОВ — Наука и искусство в цифровую эпоху: проблема синергии 116

А.Ю. СТОРОЖУК (Новосибирск) — Искусственный интеллект: как углеродная жизнь создала кремневую 134

НАУЧНАЯ ЖИЗНЬ

Обзоры, объявления, сообщения ■

С.Ф. СЕРГЕЕВ (Санкт-Петербург) — Санкт-Петербургское отделение Научного совета по методологии искусственного интеллекта и когнитивных исследований РАН в 2020 году 149

А.Ю. НЕСТЕРОВ (Самара), А.В. НИКОНОРОВ (Самара), А.В. КУПРИЯНОВ (Самара) — О деятельности Самарского отделения Научного совета по методологии искусственного интеллекта и когнитивных исследований РАН 155

CONTENTS

PROSPECTS FOR MANKIND. PHILOSOPHY OF HUMANITARIAN AND TECHNOLOGICAL DEVELOPMENT

Philosophy of Artificial Intelligence ■

V.A. LEKTORSKY — On the Philosophical Issues of Artificial Intelligence and Cognitive Studies 7

D.I. DUBROVSKY — The Task of the Creation of Artificial General Intelligence and the Problem of Consciousness 13

T.S. AKHROMEYEVA, G.G. MALINETSKIY — New Pythagoreanism, Complexity, Artificial Intelligence 45

E.A. BEZLEPKIN (Novosibirsk), A.S. ZAYKOVA (Novosibirsk) — Neurophilosophy, Philosophy of Neuroscience, and Philosophy of Artificial Intelligence: The Problem of Distinguishing 71

Development Trends of Artificial Intelligence ■

V.E. LEPSKIY — Artificial Intelligence in Subject-Oriented Control Paradigms 88

A.V. SHILLER, O.E. PETRUNYA — Architectural Approach to Design of Emotional Intelligent Systems 102

COGNITIVE SPACE

Digital Culture: Problems and Prospects ■

V.G. BUDANOV, A.R. EFIMOV — Science and Art in the Digital Age: A Synergy Problem 116

A.Yu. STOROZHUK (Novosibirsk) — Artificial Intelligence: How Carbon-Based Life Has Created Silicon-Based Life 134

SCIENTIFIC LIFE

Reviews, Announcements, and Reports ■

S.F. SERGEEV (Saint Petersburg) — Saint Petersburg Branch of the RAS Scientific Council on the Methodology of Artificial Intelligence and Cognitive Research in 2020 149

A.Yu. NESTEROV (Samara), A.V. NIKONOROV (Samara), A.V. KUPRIYANOV (Samara) — On the Work of the Samara Branch of the RAS Scientific Council on the Methodology of Artificial Intelligence and Cognitive Research 155

ПЕРСПЕКТИВЫ ЧЕЛОВЕЧЕСТВА. ФИЛОСОФИЯ ГУМАНИТАРНО-ТЕХНОЛОГИЧЕСКОГО РАЗВИТИЯ

Философия искусственного интеллекта

DOI: 10.30727/0235-1188-2021-64-1-7-12
Введение в рубрику
Section introduction

О философских проблемах искусственного интеллекта и когнитивных исследований

В.А. Лекторский
Институт философии РАН, Москва, Россия

Лекторский Владислав Александрович – доктор философских наук, академик РАН, главный научный сотрудник сектора теории познания Института философии РАН.
v.a.lektorski@gmail.com
https://orcid.org/0000-0001-8496-7814

Для цитирования: *Лекторский В.А.* О философских проблемах искусственного интеллекта и когнитивных исследований // Философские науки. 2021. Т. 64. № 1. С. 7–12. DOI: 10.30727/0235-1188-2021-64-1-7-12

On the Philosophical Issues of Artificial Intelligence and Cognitive Studies

V.A. Lektorsky
Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia

Vladislav A. Lektorsky – D.Sc. in Philosophy, Full Member of the Russian Academy of Sciences, Chief Research Fellow, Department of the Theory of Knowledge, Institute of Philosophy, Russian Academy of Sciences.
v.a.lektorski@gmail.com
https://orcid.org/0000-0001-8496-7814

For citation: Lektorsky V.A. (2021) On the Philosophical Issues of Artificial Intelligence and Cognitive Studies. *Russian Journal of*

Philosophical Sciences = Filosofskie nauki. Vol. 64, no. 1, pp. 7–12. DOI: 10.30727/0235-1188-2021-64-1-7-12

В январе 2005 года основан Совет по методологии искусственного интеллекта при Отделении общественных наук Российской академии наук, впоследствии переименованный в Научный совет по методологии искусственного интеллекта и когнитивных исследований при Президиуме РАН.

Вначале я занимал должность заместителя председателя этого Совета, в настоящее время являюсь председателем. В первые годы деятельности нашего Совета многие смотрели на нас как на чудаков, занявшихся экзотической, малоперспективной и далекой от практики проблематикой. Сегодня становится очевидным, что вопросы, которым посвящена работа Совета, – это передовой край развития всех наук, как естественных, так и социально-гуманитарных. Вместе с тем речь идет о принципиально новых, интеллектуальных технологиях, в корне изменяющих экономику, социальную жизнь, человека. Когда с высокой трибуны было сказано о том, что тот, кто оседлает искусственный интеллект, будет определять мировые процессы, повсюду стали говорить о значимости разработки проблематики интеллектуальных систем, когнитивных наук и цифровизации.

За 16 лет интенсивной работы Совета создано несколько десятков отделений в России, десятки регулярно заседающих секций, охватывающих специальную проблематику исследований. С нами взаимодействуют разработчики искусственных интеллектуальных систем, математики, психологи, нейрофизиологи, лингвисты. Важнейшую роль при этом играют философы. В указанный период проведено 100 заседаний семинара Совета, несколько сотен заседаний секций. В процессе и по итогам работы Совета издано несколько книг, опубликовано множество статей.

В контексте изложенного хотелось бы высказать ряд соображений о философско-методологическом смысле проблематики искусственного интеллекта и когнитивных исследований. 70 лет тому назад английский математик А. Тьюринг написал статью «Вычислительные машины и разум», которая была

размещена на страницах ведущего британского философского журнала «Mind». Это не случайно. Текст Тьюринга построен на его математических идеях, однако данная статья имеет прежде всего философский смысл. Речь идет о том, можно ли допустить ситуацию, когда искусственный интеллект достигнет такой степени развития, что с помощью некоторых текстов невозможно будет отличить, с кем мы имеем дело – с живым человеком, т.е. естественным существом, или с искусственным интеллектом. Статья Тьюринга породила массу продолжающихся до сих пор философских дискуссий: можно ли искусственно имитировать все познавательные процессы человека, возможно ли машинное творчество, может ли искусственная интеллектуальная система понимать другого и саму себя и др. Интересно, что Тьюринг сформулировал эти вопросы, когда еще не было компьютеров. Но именно он заложил фундамент компьютерных наук, а также выявил многое из философской проблематики, которая обсуждается ныне в связи с разработками в области искусственного интеллекта.

После публикации Тьюринга появились первые компьютеры, а спустя некоторое время возникла новая дисциплина – когнитивная наука. Это междисциплинарное движение включало в себя разработчиков искусственных интеллектуальных систем, философов, специалистов в области когнитивной психологии, нейронауки, когнитивной лингвистики. Центральная идея когнитивной науки заключалась в том, что наше сознание, мышление работает по типу гигантского компьютера. Согласно мнению основателей когнитивной науки, человек – несовершенная интеллектуальная система, а искусственный интеллект может стать совершенной системой такого рода.

Со временем в системе когнитивных исследований произошли существенные изменения. Сегодня приоритетными в этих разработках стали т.н. когнитивные нейронауки – попытки понять работу нашего мозга с помощью вычислительных моделей. В итоге получены огромные результаты, сделаны интересные открытия (например, зеркальных нейронов), предложены различные модели работы сознания. Большинство ис-

следователей в этой области, если и не разрабатывают модели искусственного интеллекта, то пытаются их применять.

В качестве наиболее важного аспекта исследований в области искусственного интеллекта выступает не просто понимание сознания, мышления, человека, а мысль о том, что искусственные интеллектуальные системы всё активнее входят в нашу жизнь. Это – уже реальность, а не проект или тем более фантазия. И эта реальность изменяется катастрофически быстро. В этой связи возникает много новых проблем, начиная от экономических, социальных, оборонных, политических и заканчивая вопросами о том, каково место человека в новой реальности, которую не так давно называли «электронным обществом», а сегодня – «цифровым». Необходимо учитывать, что цифровизация основана на использовании искусственных интеллектуальных систем, без которых она была бы невозможна. Цифровизация – наша жизнь, и в этой реальности мы существуем.

Кажется, что на пути цифровизации человечеству под силу реализовать то, о чем люди мечтали на протяжении столетий. Однако в этой связи возникает и множество вопросов, решение которых не всегда ясно. Например, в условиях цифровизации слишком удешевляется производство товаров, и ряд специалистов даже говорят о возможности того, что раньше считалось немыслимым: о возникновении ситуации, когда некоторые товары будут раздавать бесплатно. При глобальной цифровизации изменяется рынок как таковой, поскольку можно просчитывать результаты, которые раньше просчитать было нельзя. Ряд привычных профессий исчезнут. Распространение социальных сетей изменяет границу приватного и публичного пространства, что ставит под вопрос привычные представления об индивидуальной автономии и идентичности. Появляется возможность онлайн-образования. Но как мы смогли убедиться в условиях пандемии коронавируса, обучение в режиме онлайн не заменяет личного общения педагога с учащимися в реальном, а не виртуальном мире. Как показывают исследования, чтение лишь электронных учебников («цифровое чтение») и коммуникация через интернет не дают полноценного образования.

Но главный вопрос состоит в том, что происходит с самим человеком. Великий мыслитель И. Кант считал, что философия отвечает на три ключевых вопроса. Что я могу знать? Что я должен делать? На что я могу надеяться? Ответить на них означает ответить на главный вопрос о том, что такое человек. Данные вопросы требуют новых ответов. Попытаемся ответить на каждый из них.

Что я могу знать? По мнению адептов и фанатиков глобальной цифровизации, можно знать только то, что можно просчитать. С этой точки зрения полноценным знанием о чем-то могут обладать только интеллектуальные технические устройства, умные машины. Именно на такую машину нужно теперь полагаться даже в простейших жизненных ситуациях. Между тем индивидуальная идентичность конструируется с помощью знания о своем прошлом. Но происходящая цифровизация памяти создает принципиально новые возможности для вмешательства в нашу жизнь. Необходимо помнить о том, что наши действия оставляют цифровые следы, которые доступны внешним инстанциям, способным нас контролировать. В итоге мы перестаем быть владельцем информации о собственной жизни и ее хозяином. Личное пространство любого может быть взломанным, и в этом случае можно стать предметом управления со стороны других людей.

Что я должен делать в эпоху глобальной цифровизации? Очевидно, что в условиях, когда происходит утрата индивидуальной автономии, не мы будем решать, что нужно делать: умные технические устройства будут предписывать нам эти решения. Человек, который считал, что с помощью искусственного интеллекта он будет управлять миром, на самом деле становится управляемым с помощью этого интеллекта. Тем самым такой человек превращается из Бога, демиурга Вселенной, в раба, поскольку искусственный интеллект лучше знает, что нужно такому человеку.

На что я могу надеяться? Вроде бы надеяться и не на что, хотя надежду подают некоторые футурологи. В частности, на сайте Российского футурологического общества размещен прогноз о том, что к середине XXI века возникнет цифровое бессмертие,

т.е. человек получит возможность существовать вечно в виде электронного двойника. В современном мире немало людей (физики, нейроученые, психологи, философы, футурологи) верят в подобную возможность и пытаются вести практическую работу в данном направлении. Такая идея представляется мне сомнительной.

В ряде работ я обосновал мысль о том, что, если бы человек стал бессмертным, он перестал бы быть человеком. И в этом обществе потеряли бы смысл особенно значимые ценности, делающие человека человеком: забота о ближних, самопожертвование, мужество, сострадание. Утрата смысла смерти ведет к утрате смысла жизни. И не следует забывать о том, что даже если бы удалось осуществить «цифровое бессмертие», о чем мечтают некоторые футурологи, в ситуации глобальной цифровизации жизни человек в любом случае оказался бы управляемым и контролируемым сверхмощным искусственным интеллектом. Если бы при этом у него осталось что-то человеческое, вряд ли он захотел бы такого бессмертия.

Человечество стоит перед лицом экзистенциальных вызовов. Один из них связан с глобальной цифровизацией на основе использования искусственного интеллекта и разработок в области когнитивных наук. Цифровизация приносит немало благ, решает ряд проблем, но она способна создать и угрозу в отношении бытия человека. Это означает, что предлагаемые в рамках цифровизации проекты нуждаются в философской экспертизе, направленной на гуманизацию информационных технологий и культивирование высших человеческих ценностей: свободы, личной автономии, достоинства, идентичности, творчества, понимания, взаимопонимания. Цифровизация может и должна вывести человека на новый уровень творческой жизни, а не привести его к самоуничтожению. Философия бывает максимально востребована в периоды культурных и познавательных кризисов. Мы живем именно в такое время. Сегодня время философии.

DOI: 10.30727/0235-1188-2021-64-1-13-44
Оригинальная исследовательская статья
Original research paper

Задача создания Общего искусственного интеллекта и проблема сознания*

Д.И. Дубровский
Институт философии РАН, Москва, Россия

Аннотация

В статье рассмотрена задача создания Общего искусственного интеллекта – системы искусственного интеллекта, которая приближается к функциональным возможностям естественного интеллекта. Сделан акцент на ведущей роли Сбербанка в актуализации и организации программы исследований в целях разработки этой проблемы, имеющей для России стратегическое значение. Успешная реализация данной программы предполагает решение принципиальных методологических вопросов, требующих участия философов – специалистов в области эпистемологии и методологии науки. Показано соотношение понятий «искусственного и «естественного» «сильного» и «слабого», общего» и «узкого» интеллекта. Выявлены теоретические трудности, связанные с четким определением свойств Общего интеллекта и способов его практической реализации. Автор статьи обращает внимание на значение исследований сознания для разработки Общего искусственного интеллекта и приходит к выводу о том, что на первом плане находятся вопросы использования результатов феноменологического анализа субъективной реальности, ее ценностно-смысловых и операциональных структур. Такие вопросы освещены подробно, поскольку они имеют прямое отношение для построения новых когнитивных архитектур. Последние позволяют выйти за пределы «узкого» искусственного интеллекта и создать Общий искусственный интеллект, способный обладать высокой степенью автономности и самостоятельного решения широкого комплекса задач в разных условиях внешней среды. В этой связи показана ограниченность операционалистской методологии Тьюринга, которая исключает применение результатов специальных исследований сознания как субъективной реальности. Использование таких результатов связано с разработкой посттьюрин-

* Работа выполнена при поддержке РНФ, проект № 21-18-00184 «Социогуманитарные основания критериев оценки инноваций, использующих цифровые технологии и искусственный интеллект».

говой методологии, открывающей значительные возможности для создания Общего искусственного интеллекта.

Ключевые слова: естественный интеллект, искусственный интеллект, общий интеллект, узкий интеллект, мышление, сознание, субъективная реальность, операциональные динамические структуры субъективной реальности, факторы самоорганизации.

Дубровский Давид Израилевич – доктор философских наук, профессор, главный научный сотрудник Института философии РАН.
ddi29@mail.ru
https://orcid.org/0000-0003-4392-2526

Для цитирования: *Дубровский Д.И.* (2021) Задача создания Общего искусственного интеллекта и проблема сознания // Философские науки. 2021. Т. 64. № 1. С. 13–44. DOI: 10.30727/0235-1188-2021-64-1-13-44

The Task of the Creation of Artificial General Intelligence and the Problem of Consciousness*

D.I. Dubrovsky
Institute of Philosophy, Russian Academy of Science, Moscow, Russia

Abstract

The article discusses the task of creation of Artificial General Intelligence (AGI), that is, an artificial intelligence system that approaches the functional capabilities of natural intelligence. Emphasis is made on the leading role of Sberbank in updating and organizing a special research program in order to develop this problem, which is of strategic importance for Russia. It has been established that the successful implementation of this program presupposes solving fundamental methodological issues that require input of philosophers – specialists in the field of epistemology and methodology of science. We show the relations of the concepts of "artificial" and "natural," "strong" and "weak," "general" and "narrow" intelligence. The article reveals the theoretical difficulties associated with a clear definition of the properties of general intelligence and the ways of its practical implementation. The author draws attention to the importance of studies of consciousness for the development of general artificial intelligence and comes to the conclusion that the priority issues are using the results of the phenomenological analysis of subjective reality, its value-semantic and op-

* This work was supported by the Russian Science Foundation (RSF), project no. 21-18-00184 "Social and Humanitarian Foundations of Criteria for Evaluating Innovations Using Digital Technologies and Artificial Intelligence."

erational structures. The paper discusses these issues in detail, as they are of direct relevance to the construction of new cognitive architectures. The latter make it possible, to go beyond the limits of "narrow" artificial intelligence and create AGI with a high degree of autonomy and independent solutions to a wide range of problems in different environments. It demonstrates the limitation of Turing's operationalist methodology that excludes the use of the results of special studies of consciousness as a subjective reality. Application of such results is associated with the development of post-Turing methodology, which opens up significant opportunities for creating an AGI.

Keywords: natural intelligence, artificial intelligence, general intelligence, narrow intelligence, thought, consciousness, subjective reality, operational dynamic structures of subjective reality, factors of self-organization.

David I. Dubrovsky – D.Sc. in Philosophy, Professor, Chief Research Fellow, Department of the Theory of Knowledge, Institute of Philosophy, Russian Academy of Sciences.
ddi29@mail.ru
https://orcid.org/0000-0003-4392-2526

For citation: Dubrovsky D.I. (2021) The Task of the Creation of Artificial General Intelligence and the Problem of Consciousness. R*ussian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 13–44.
DOI: 10.30727/0235-1188-2021-64-1-13-44

Введение

Основные цели и задачи развития в нашей стране искусственного интеллекта (далее – ИИ) четко сформулированы в «Национальной стратегии развития искусственного интеллекта на период до 2030 года», утвержденной Указом Президента Российской Федерации от 10 октября 2019 года. В настоящее время развитие ИИ подошло к новому стратегическому рубежу, т.е. к задаче создания **Общего ИИ** (в принятой международной номенклатуре – **AGI**, **Artificial General Intelligence**). Эта задача соответствует мировому тренду развития ИИ в условиях нарастающей конкуренции между крупнейшими научными центрами и специализированными корпорациями, – а более широком масштабе, – между Россией и такими лидирующими в области ИИ государствами, как США, Китай, Япония, страны Западной Европы. Факт наличия острой конкуренции очевиден,

поскольку Россия пока существенно отстает от своих главных конкурентов. Догонять их по основным сложившимся приоритетным направлениям развития ИИ – бессмысленно.

Нам нужен прорыв! Прорыв на принципиально новый уровень развития ИИ, о чем говорил Президент России В.В. Путин. Для всех является понятным, в чем заключается стратегическое значение прорыва, от которого зависит наше экономическое развитие, решение проблем безопасности, многих важных социальных вопросов. Но возможно ли это? Теоретически прорыв (несмотря на наше положение в настоящее время) возможен с учетом того, что проблемное поле ИИ в высшей степени многомерно. В нем назревает сегодня множество мыслимых и немыслимых пока новаций разного рода и разного порядка. Вопрос состоит в том, как своевременно диагностировать и использовать или создать такие из них, которые позволили бы добиться выхода на принципиально новый уровень разработки Общего интеллекта (AGI). Наша страна обладает высокоразвитой научно-технологической базой, сильной математической школой для подобных исследований, располагает рядом крупных достижений и преимуществ в решении теоретических, организационных и практических проблем развития ИИ, которые могут быть успешно использованы в нынешней конкурентной ситуации.

Фундаментальный прорыв означает формирование новой теоретической и методологической основы ИИ, предусматривает разработку новых подходов и методов, которые способствовали бы созданию AGI, т.е. системы, приближающейся по функциональным возможностям к естественному интеллекту (далее – ЕИ).

О соотношении понятий общего и узкого, сильного и слабого, естественного и искусственного интеллекта

В актуализации, организации, стимулировании исследований в области AGI ведущую роль сыграла активная позиция Сбербанка России, которому удалось создать за счет собственных средств мощную научно-исследовательскую команду и вычислительную инфраструктуру в области ИИ и робототехники. В рамках этой структуры, объединяющей ведущих специали-

стов, созданы экспертные советы, на которых подробно обсуждают ключевые проблемы ИИ и AGI, ежегодно проводится одна из крупнейших мировых конференций по искусственному интеллекту «Artificial Intelligence Journey» (8–9 ноября 2019 года, 4–5 декабря 2020 года), где постоянно выступает Президент России В.В. Путин.

В 2021 году вышла первая в нашей стране книга, посвященная основательному обзору проблематики AGI, анализу состояния ее разработки, основных трудностей и возможных способов их преодоления. В предисловии к ней президент, председатель правления Сбербанка России Герман Греф оценивает современное состояние разработок ИИ: «В большинстве своем существующие решения являются примерами реализации технологий узко специализированного искусственного интеллекта, требующего настройки и перепроверки со стороны человека. Чтобы так же хорошо решать разнообразные комплексные задачи, как это делают люди, машины должны научиться строить причинно-следственные модели окружающей среды и ориентироваться в разных контекстах, а не просто максимизировать успех при решении какой-то узкой задачи. Они должны понимать физические, психологические и другие законы нашего мира и уметь связывать новую информацию в общую картину с тем, что уже знают. Чтобы добиться этого, нам необходимо преодолеть очередной технологический рубеж – создание Общего искусственного интеллекта или AGI» [Сильный ИИ… 2021, 5–6]. Греф подчеркивает принципиальное значение разработки в этих целях междисциплинарных проблем AGI и формулирует стратегическую установку для достижения успеха: «Крайне важным фактором создания и развития технологий Общего искусственного интеллекта является обеспечение сквозного целеполагания между прикладными (или индустриальными) задачами, фундаментальными исследованиями и системой образования – так называемая триада Practice – Education – Research» [Сильный ИИ… 2021, 7]. Указанные три составляющие ввиду их неотъемлемой взаимообусловленности должны быть в центре внимания государственных органов, призванных содействовать реализации программы AGI.

Далее рассмотрим теоретико-методологические вопросы философского плана. Для создания AGI в первую очередь следует преодолеть ограниченности традиционной методологии разработки ИИ, основы которой заложены А. Тьюрингом. Согласно этой методологии, понятие «интеллект» трактуется в сугубо функционально-операциональном смысле, исключающем роль сознания, вся сложная проблематика которого выносится за скобки. Исследование сознания представляется излишним в данном случае: существует когнитивная или практическая задача, которая формулируется и решается посредством сугубо операциональных методов с помощью компьютерных программ.

Именно методология Тьюринга открыла компьютерную эру, обеспечила выдающиеся успехи ИИ, развитие информационных технологий, качественное изменение земной цивилизации. Однако сегодня мы переходим к новому этапу, нуждающемуся в существенных теоретико-методологических обновлениях. Парадигма функционализма, в формирование которой А. Тьюрингом внесен первостепенный вклад, естественно, остается действующей, но она требует более широкой интерпретации с учетом ее роли в объяснении сознания и способов использования новейших результатов исследований сознания для развития ИИ.

Тенденция «реабилитации» сознания и потребность коррективов методологии Тьюринга отчетливо проявлялась в разных формах за последние два десятилетия. Но в этом направлении нам предстоит глубокая теоретико-методологическая работа. К тому же немало специалистов в области ИИ опираются на традиционные методы, позволяющие и в настоящее время успешно решать многие практические задачи, и убеждены, что для реализации намечаемых больших задач достаточно совершенствования, улучшения существующих методов. Они скептически относятся к привлечению проблемы сознания, считая, что она влечет за собой множество неопределенностей, «сплошной туман», что нет смысла в разработке т.н. трудной проблемы сознания, только еще более усложняющей ситуацию. Подобные убеждения проявились у ряда некоторых видных

специалистов, выступавших на экспертных обсуждениях проекта разработки AGI, предложенного Сбербанком России. Соответстующее мнение сложилось у меня, поскольку я участвовал в этих обсуждения полемизировал по этим вопросам.

Соответствующие настроения присущи, конечно, лишь некоторым специалистам. Существуют разные взгляды, разные направления разработки ИИ. Среди них – концептуальные подходы, в которых внимание обращено на изучение ЕИ, моделирование его специфических свойств, создание новых когнитивных архитектур с целью сближения функциональных возможностей ИИ и ЕИ. Одним из тех, кто считает ЕИ главным источником и ресурсом нового этапа развития ИИ (что отвечает задачам AGI), является крупный отечественный специалист В.К. Финн. Он всесторонне анализирует функциональные когнитивные структуры ЕИ и, главное, придает огромное значение философско-методологическому подходу к решению актуальных проблем современного этапа развития ИИ, использует его в своих исследованиях [Финн 2011]. В.К. Финном создан широко известный ДСМ-метод, на основе которого автором получены значительные результаты [Финн 2016]. Развитие этого направления может быть перспективным и для разработки AGI. Существуют и многие другие подходы, применяющие в этом аспекте результаты исследований ЕИ (см., например: [Кузнецов 2016]).

Несомненно, традиционные методы с учетом их совершенствования еще долго будут успешно использоваться для решения типичных прикладных задач. Но для прорыва нужны колоссальные творческие и волевые усилия, связанные с преодолением привычных ментальных клише, выбором (и созданием) наиболее перспективных направлений исследований, концентрацией сил и средств для разработки принципиально новых теоретических и методологических концепций развития ИИ, новых подходов к решению междисциплинарных проблем.

В фокусе внимания исследователей оказываются проблемы теории, методологии, междисциплинарности, понятия интеллекта, мышления, познания и сознания, проблемы и гипотезы,

объяснения и понимания, вопрос о соотношении эмпирического и теоретического знания, характера связи между ИИ и ЕИ. Первостепенная роль в их осмыслении принадлежит философам – специалистам в области эпистемологии (теории познания), методологии науки, феноменологии, изучения сознания в широком смысле включая те направления, которые в западной номенклатуре именуются аналитической философией сознания.

Однако нужен *особый вид специализации философа*, работающего на стыке философии и науки, требующий глубоких методологических знаний и теоретических навыков, аналитического опыта и, конечно, достаточной компетенции в соответствующей области научного знания. В связи с этим центральное место занимают вопросы о понятии интеллекта и соотношении всех его специальных модификаций, связанных с понятиями ИИ и ЕИ: сильный и слабый, общий (AGI) и узкий ИИ, особенно *соотношение AGI с ЕИ*. К этому добавляются вопросы о прото-AGI и узкого AGI, которые рассматриваются как стадии продвижения к полноценному AGI. Все это получило широкое освещение в указанной книге [Сильный ИИ 2021], которая, подчеркнем еще раз, вносит исключительно большой вклад в разработку AGI.

Под сильным интеллектом понимают человеческий интеллект, наделенный сознанием, т.е. ЕИ. Все существующие формы ИИ чаще относят к слабому интеллекту. Но все, что касается Общего интеллекта (AGI) и его отношения к ЕИ, к обычному ИИ, сопровождается вопросами, ответы на которые истолкованы неоднозначно. Термин «Общий интеллект» появился в 1997 году, когда компьютер *Deep Blue* обыграл чемпиона мира в шахматы. На базе понятия общего интеллекта было предложено делить ИИ не на сильный и слабый, а на *общий* (способный адаптироваться к решению разных проблем) и *узкий* (способный решать определенную задачу). В 2007 году состоялась первая конференция, посвященная проблемам развития Общего ИИ, которые впоследствии стали предметом возрастающего внимания специалистов в области ИИ. Тем более что подавляющее большинство систем ИИ характеризуются как узкоспециализи-

рованные, способные качественно решать определенную узкую задачу. Поэтому критики не раз говорили о том, что компьютер, запрограммированный на решение тысячи задач, не способен самостоятельно научиться решать тысячу первую.

В указанной нами книге о сильном ИИ подробно рассказано об истории развития ИИ, ее значимых этапах, о том, что такое развитие происходило в направлении снижения зависимости технологий ИИ от человека, их способности к решению более широкого круга задач. Но достижения по-прежнему далеки от решения проблемы AGI. Авторами издания дана общая оценка нынешнего положения дел:

«Во-первых, с точки зрения адаптивности все существующие системы ИИ на основе любых известных подходов предполагают функционирование в ограниченном наборе задач и условий и не способны самообучаться функционированию в условиях существенно новых. То есть они являются программируемыми, хотя сложность их программирования заметно снизилась, а способность к обучению возросла. Во-вторых, с точки зрения автономности систем ИИ все существующие системы не являются автономными и не могут полноценно функционировать без живого оператора, отвечающего за запуск и остановку, техобслуживание, целеполагание и определение режимов работы в зависимости от тех или иных условий или задач. То есть они остаются управляемыми» [Сильный ИИ… 2021, 52]. Авторы книги продолжают: «В-третьих, с точки зрения интегративности современные системы ИИ являются не системами, обладающими интеллектом как таковым (даже ограниченным), а системами компьютерного зрения, обработки естественного языка, анализа данных (машинного обучения), обработки символьной информации (рассуждений на основе знаний) и т.д., то есть интегративными не являются» [Сильный ИИ… 2021, 52]. Тем самым четко обозначены главные, пока отсутствующие свойства, требуемые для построения действительного AGI.

Интерес представляет то, как решен в книге вопрос об определении AGI, его степени общности и отношении к ЕИ. Авторы анализируемого издания ссылаются на ряд определений: «Общий интеллект – это способность достигать

сложных целей в сложных средах» (Бен Герцель). «Интеллект – это способность системы адаптироваться к своей среде, работая при недостаточных знаниях и ресурсах» (Пей Ванг) [Сильный ИИ… 2021, 28]. Находит отражение и еще одно утверждение: «Делая акцент на широком диапазоне сред, область AGI позволяет нам избавиться от антропоцентричных предпочтений и предлагает сфокусироваться на общих решениях, пригодных для разных агентов (человека, животных, роботов, ботов и т.д.), действующих в разных условиях…», в том числе «на неизведанных планетах» [Сильный ИИ… 2021, 29].

Как видим, широкая трактовка интеллекта теоретически правомерна, соответствует принципу изофункционализма систем и принципу инвариантности информации по отношению к физическим свойствам ее носителя (т.е. возможности одной и той же информации кодироваться разными способами, иметь различное физическое воплощение). Не исключено, что такая трактовка интеллекта может быть полезной за пределами не только антропологического, но и биологического контекста, оставляя простор для творчества технологических суперноваций, для осмысления гибридных и иных новообразований в ходе идущей нарастающими темпами антропотехнологической эволюции.

Однако возникает вопрос о том, почему сразу отодвигается на задний план «сильный» интеллект, наш ЕИ как высшая форма из всех известных форм интеллекта, наделенная сознанием, благодаря которому мы создаем все искусственные интеллекты. Ведь именно в ЕИ присутствуют драгоценные и желательные для AGI свойства.

Авторы разъясняют свою позицию следующим образом: «Определение интеллекта в области AGI может казаться слишком абстрактным и далеким от наших представлений о естественном интеллекте. А еще широко распространено мнение, что человеческий интеллект – единственный пример интеллекта. Почему бы не опираться на него? Даже если мы предположим, что для характеристики AGI необходимы или достаточны критерии человеческого интеллекта, описание последнего тоже основано на довольно зыбких понятиях»

[Сильный ИИ… 2021, 30]. В подтверждение о «зыбкости» дано определение интеллекта из Википедии, далеко не надежного в научном отношении источника. В нем, наряду с указанием на то, что он является «качеством психики», перечислены в несистематизированном виде его многочисленные «способности», от приспособления к новым ситуациям до ощущений, мышления, воображения, внимания, воли и рефлексии.

Приведенное определение трудно принять в полной мере, поскольку непонятно, каким образом упорядочить, соотнести между собой эти способности и использовать их для целей AGI. В итоге предложено не вполне ясное заключение: «Определение интеллекта в области AGI не включает эти способности, но и не отвергает их» [Сильный ИИ… 2021, 31]. Заслуживает внимания и весомый аргумент: «Как повышение уровня решения узких задач вплоть до сверхчеловеческого не потребовало “сильных” качеств, так и расширение общности методов решения задач вовсе не обязательно подразумевает преднамеренное движение в сторону сильного ИИ» [Сильный ИИ… 2021, 32]. Под «сверхчеловеческим» имеются в виду действительные достижения «узкого» ИИ: например, победа компьютера над чемпионом мира по игре в го, многие другие результаты, связанные с задачами анализа больших данных, быстродействия, использования ИИ в производственных и коммуникативных процессах.

Но далее читаем:

> Можно предположить, что *некоторые аналоги некоторых «сильных» качеств у действительно общего ИИ должны быть* (здесь и далее курсив и выделение полужирным мои. – *Д. Д.*). Например, наверняка общий ИИ должен иметь ***способность к интроспекции*** – анализу ***собственных мыслительных процессов*** или даже оптимизации лежащих в их основе алгоритмов. При этом у такого ИИ будет некий ***образ себя как часть картины мира***. Но это не обязательно означает, что у него будет самосознание в философском смысле и уж тем более личность сродни человеческой, хотя по глубине рефлексии он вполне может и превосходить человека. Наверняка он проявит «понимание» тех областей, в которых действует успешнее человека. Но это не значит, что такое понимание будет сопровождаться у него субъективными переживаниями, схожими с человеческими. Наверняка у него будет многомерная система мотивации, включающая аналоги, например, любопытства и удивле-

ния. Но его вовсе не обязательно пытаться наделить всеми человеческими эмоциями. Хотя, скажем, для социальных роботов это может быть полезно, но даже они способны лишь симулировать чувства и эмоции, а не испытывать их [Сильный ИИ… 2021, 32–33].

Приведенные суждения показательны. Они свидетельствуют о том, что специфические свойства и функции сознания, присущие нашему мыслительному процессу, должны привлекаться для построения AGI. Но возможно ли это за вычетом состояния ***субъективной реальности – неотъемлемого качества сознания?***

Теоретически возможно, даже необходимо, но при условии адекватного вычленения и описания оперативных аналогов определенных состояний субъективной реальности. Что имеется в виду под «интроспекцией» и «собственными мыслительными процессами» или «образом себя как части картины мира»? Эти понятия выражают специфические особенности реальной сознательной деятельности (исследуемого нами ЕИ). Они являются весьма сложными, многомерными, взаимодополняющими друг друга по содержанию. В когнитивном плане каждое из них включает в себя возможность множества аналитических и синтетических оперативных действий. Поэтому не вполне ясно, о каком аналоге идет речь. Приведенные понятия представлены слишком абстрактно. Следовательно, указания на подобные аналогии вряд ли могут иметь предполагаемое значение для моделирования когнитивных структур AGI.

Несмотря на это, установка на поиск аналогий заслуживает поддержки, т.к. обращена к задаче использования специфических свойств сознательной деятельности для разработки AGI. Нужны результаты профессионального философского и специального феноменологического анализа операциональных структур субъективной реальности, допускающих корректное определение их аналогов, пригодных для построения когнитивных архитектур AGI.

Такая установка отличается от жестких высказываний о том, что Общий искусственный интеллект «не лежит между слабым и сильным ИИ, а просто находится в стороне от них и опреде-

ляется без отсылки к человеческому интеллекту как ИИ, способный решать широкий круг задач» [Сильный ИИ… 2021, 27]. И она противостоит упомянутым выше случаям прямого отрицания роли сознания в разработке проблем ИИ и AGI. Специалист в области ИИ, занимающий эту позицию, демонстрирует парадоксальную ситуацию: он ***как бы забывает*** о том, что является носителем ЕИ и обладает искомыми свойствами AGI, постоянно применяет их в своей деятельности, несмотря на то, что они скрыты для него в хитросплетениях собственного сознания.

Большинство специалистов в области ИИ занимаются решением прикладных задач. Они не могут быть компетентны в философско-методологических вопросах в контексте проблем сознания, анализа ценностно-смысловых и деятельно-волевых структур субъективной реальности; им «хватает» своих дел, трудностей и проблем. Но сегодня важно понимать, что профессионалы, занимающиеся ***теорией познания, методологией науки***, ***феноменологией субъективной реальности***, способны быть активными помощниками в общем деле, в частности в вопросах теоретического осмысления роли сознания на современном этапе развития ИИ. Разумеется, многие специалисты в области ИИ отдают себе в этом отчет и готовы к сотрудничеству.

Сознание и AGI. Субъективная реальность: ее общие свойства, ценностно-смысловые и интенционально-волевые оперативные структуры

Исследования сознания, взятого в его специфическом и неотъемлемом качестве субъективной реальности, позволяют сформулировать основные системные, структурные, функциональные, операциональные характеристики естественной мыслительной деятельности человека, а значит, представляют собой важный ресурс для моделирования когнитивных структур, отвечающих задачам построения AGI. С этой целью актуальным становится дальнейшее рассмотрение основных философских и научных положений о сущности самого качества субъективной реальности, ранее изложенных (эти вопросы

подробно освещались в ряде моих работ: [Дубровский 2002; Дубровский 2020].

Качество нередко выносится за скобки теми, кто занимается задачами сугубо функционалистского, бихевиорального типа (как «сделать» что-то, в том числе создать программу для компьютера или создать робота). В большинстве описанных случаев нет необходимости выделять и анализировать качество субъективной реальности, поскольку внимание сосредоточено на поставленной задаче и наличном процессе деятельности; мы будто не замечаем это качество, и даже когда напрягаем, корректируем собственное мышление; оно подобно воздуху, который мы тоже «не замечаем» во многих интервалах жизни. Качество субъективной реальности становится слишком «заметным», когда предметом исследования оказывается мышление и сознание. При этом возникает целый ряд принципиальных вопросов. Рассмотрим каждый из них.

1. Именно качество субъективной реальности создает главные теоретические трудности проблемы сознания. Они связаны в первую очередь с тем, что явлениям субъективной реальности нельзя приписывать физические свойства (массу, энергию, пространственные характеристики). Возникает вопрос о том, как, не обращаясь к идеалистическим постулатам о наличии особой духовной субстанции, построить теоретически корректное объяснение сознания с материалистических позиций, учитывая его очевидную связь с физическими процессами. Вопросы, относящиеся к требуемому объяснению, принято объединять под общим названием «Трудная проблема сознания». Она более 60 лет находится в центре внимания западной аналитической философии, служит предметом острых дискуссий. Ей посвящены тысячи публикаций. Однако общепринятое решение отсутствует. В большей части концепций предлагается тот или иной редукционистский подход, чаще всего в форме отождествления явлений субъективной реальности с физическими процессами (см. подробный критический анализ этих концепций [Дубровский 2019]). В отличие от последних мною предложен нередукционистский способ решения глав-

ных теоретических вопросов «Трудной проблемы сознания» [Дубровский 1971; Дубровский 2015; Dubrovsky 2019b], который может представлять определенный интерес и для разработки ряда вопросов ИИ.

«Трудная проблема сознания» рассматривается в двух главных категориальных планах: ***онтологическом*** и ***эпистемологическом:***

1) в ***онтологическом плане*** поставлены вопросы объяснения связи явлений субъективной реальности с деятельностью мозга [Koch, Massimini, Boly, Tononi 2016], их способности управлять телесными процессами, а также более широкие вопросы происхождения субъективной реальности, форм ее существования, роли в физическом мире (при условии, что ей нельзя приписывать физические свойства);

2) в ***эпистемологическом плане*** появляется проблема перехода от индивидуально-субъективного опыта, присущего только данному индивиду (выражаемого в отчетах от первого лица), к *интерсубъективным*, общезначимым утверждениям (от третьего лица) и обоснованию истинного знания. Тем не менее все специфические для человека мыслительные процессы, связанные с решением практических и когнитивных задач, характерных и для ИИ, осуществляются в форме субъективной реальности. Поэтому апробированные результаты ее феноменологических исследований могут понадобиться для осмысления способов создания AGI, прежде всего для построения его когнитивных архитектур, способных преодолеть ограниченность когнитивных архитектур узкого ИИ.

Вместе с тем для проблематики AGI важное значение имеют вопросы онтологического плана, т.е. объяснения ***активности сознания*** (каким образом явления субъективной реальности могут служить причиной телесных, физических изменений, как объяснить способность целеполагания, свободы воли, способность моего Я управлять моими физическими действиями). Ведь для систем узкого ИИ характерно то, что они плохо работают с причинно-следственными связями; для агента же, способного действовать в широком диапазоне сред, необходимо «понимание» причинно-следственных связей, объ-

ективно присущих любой среде, в которой он должен учиться ориентироваться и действовать; для агента значимо умение различать физические причины и информационные, способность «понимать» специфику информационной причинности, видом которой служит психическая причинность. Для этого и требуется решение «Трудной проблемы сознания».

2. Понятие субъективной реальности означает всякое осознаваемое состояние, в отличие от бессознательных процессов, которые существуют одновременно с ним и служат его непременной основой. Оно охватывает как отдельные осознаваемые явления субъективной реальности и их виды (ощущения, восприятия, чувства, мысли, желания, целеполагания, волевые усилия и т.д.), так и их целостное персональное образование, объединяемое нашим Я.

Субъективная реальность *персональна*, переживается индивидом как «текущее настоящее», т.е. *сейчас*, хотя может отображать прошлое и будущее (она удостоверяет для меня реальность моего существования). Речь идет о непрерывном процессе, который временно прерывается в условиях глубокого сна без сновидений, общего наркоза, обморока, в иных случаях временной потери сознания (в этих интервалах я для себя не существую). Состояние субъективной реальности выражает актуально протекающий процесс мышления, когнитивную и деятельную активность личности. Не говоря уже о повседневной, обыденной сознательной деятельности, все новые фундаментальные идеи в науке и культуре первично возникают, оформляются и развиваются в форме явлений субъективной реальности отдельных личностей.

3. Субъективная реальность в своем специфическом качестве присуща не только сознанию человека, но и психике животных, о чем свидетельствуют опыт общения с ними и данные зоопсихологии об их ощущениях, восприятиях и свойственных высшим животным формах мышления (наличие качества субъективной реальности отчетливо подтверждается, например, известными опытами с воздействием галлюциногенов на собак).

Субъективная реальность – находка биологической эволюции. У первых одноклеточных организмов носителем информации и регулятором поведения служили химические процессы. Речь идет о *допсихическом уровне информационной реальности.* Следующим этапом ее развития стало зарождение психики, обладающей качеством субъективной реальности, возникшим у многоклеточных животных, способных активно передвигаться в постоянно изменяющейся внешней среде.

Это ознаменовало возникновение ***нового типа информационного управления***, регуляции поведения, позволило решить проблему эффективной самоорганизации многоклеточного организма. Ведь его элементами являются отдельные клетки, которые представляют собой самоорганизующиеся системы со своими жесткими программами, «отработанными» эволюцией в течение многих сотен миллионов лет. Но теперь последние должны были согласовываться с общеорганизменной программой и наоборот. Это – весьма сложная задача, решение которой предполагало нахождение оптимальной ***меры*** *централизации и автономизации* контуров управления, способной обеспечить сохранение и укрепление целостности сложной живой системы, т.е. меры централизации управления, которая не нарушает фундаментальные программы отдельных клеток, и меры автономности их функционирования, которая, наряду с кооперативными и конкурентными способами их взаимодействия между собой, не препятствует их содружественному участию в реализации программ управления целостным организмом.

Аналогичная проблема эффективной самоорганизации социальных и технических систем актуальна и для ИИ, если появляются подобные задачи взаимоотношения глобальных и локальных уровней управления в сложных системах, взаимодействия множества компьютеров (и роботов), человека и роботов, автономной системы роботов и т.п. Данная проблема решается в сложной биологической системе *гораздо* ***эффективнее***, чем в случаях взаимоотношений человека и общества, человека и технических устройств, в технических системах. Это требует внимания и изучения, когда речь идет о новых разработках в

области ИИ и интеллектуальной робототехники, связанных с аналогичными задачами эффективного управления.

Возникновение у животных субъективной реальности стало исторически первой формой ***виртуальной реальности***, открывающей по мере развития все более широкий диапазон способностей к абстрагированию, обобщению, прогнозированию, планированию, пробным действиям в виртуальном плане, иным схожим операциям (в форме «мысленных экспериментов»), повышающим приспособляемость к среде. Это качество виртуальности во многом определяет способность формирования новых навыков, приспособления к изменившейся среде, т.е., по существу, главное свойство, необходимое для AGI, – способность самостоятельно решать задачи «в широком диапазоне сред».

4. В ходе антропогенеза произошло качественное развитие субъективной реальности. Возникает сознание (как высшая форма субъективной реальности), а с ним язык. Особенностью сознания по сравнению с животной психикой является то, что психическое отображение и управление становятся объектом отображения и управления. Появляется способность, по сути, неограниченного производства информации об информации, возможность высокой степени абстрагирования, резкого расширения диапазона мысленных действий, моделирования ситуаций, проектирования, целеполагания и волеизъявления, фантазирования, творческого решения задач.

Качество субъективной реальности человека открыло принципиально новый этап развития информационных процессов. Оно представляет собой два основных свойства психической деятельности: 1) ***способность индивида иметь информацию в «чистом» виде*** – в том смысле, что ее мозговой носитель для нас элиминирован: когда я переживаю образ дерева, мне дана информация об этом предмете и отображение мной данной информации, т.е. знание о том, что именно я вижу это дерево; но я ничего не знаю, не чувствую, что при этом происходит в моем головном мозге; а в нем происходят очень сложные, многоступенчатые сетевые процессы переработки информации,

предшествующие, например, переживанию мной ощущения красного или образа бегущего ко мне человека, т.е. мне дан ***итоговый результат*** сложных информационных процессов в той ***интегральной*** форме («простой» и понятной), которая необходима для адекватного поведения и решения задач в данной ситуации; 2) вместе с тем в явлениях субъективной реальности нам дана и ***способность оперировать этой информацией в «чистом» виде по своей воле*** в широком диапазоне (контролировать собственное Я, переключать внимание, направлять движение своей мысли и концентрировать ее на определенном объекте, давать волю своему воображению, ставить определенные цели и решать мысленно разные задачи). Это означает, что наше Я является самоорганизующейся структурой, способной к эффективной саморегуляции. Возникает острый вопрос о свободе воли, который вызывает немало трудностей в нейронаучных исследованиях сознания, но может получить объяснение с позиций информационного подхода [Dubrovsky 2019a].

Именно *данность информации в «чистом» виде и способность управлять ею выражают специфические черты субъективной реальности.*

5. Динамическая структура субъективной реальности с ее многообразными способами продления и смены переживаемого контента (наличного «содержания») включает в себя функциональные регистры, т.е. единство *рефлексивного* и *арефлексивного*, *актуального* и *диспозиционального.* Рефлексивное – это длящееся *сейчас осознаваемое* переживание данного контента, которое в следующий момент «уходит» в память, становится арефлексивным (заменяется другим рефлексируемым контентом, но затем может снова «возвращаться», становится рефлексируемым). Актуальное – это рефлексивное, но взятое в его отношениях и зависимости от *диспозиционального*, которое представляет собой ценностно-смысловые структуры, хранящиеся в памяти или на более глубоком психическом, а затем и на еще более глубоком допсихическом (генетическом) уровне неосознаваемых информационных процессов. Диспозициональное в большей степени выражает устойчивые,

ценностно-смысловые образования нашей психики (усвоенные нормы культуры, убеждения, проверенные опытом знания, привычки, умения, интересы, предпочтения и т.п.), которые задают направленность актуализуемых в данном интервале желаний, мыслей и действий; в ряде случаев такая направленность становится для личности непреодолимой или труднопреодолимой (как, например, при наркотической зависимости).

В аспекте рассмотрения регистров рефлексивное и арефлексивное, актуальное и –диспозициональное для разработки ИИ представляет интерес работа нашей памяти в связи с тем, что компьютерам присуща архитектура фон Неймана, основанная на постоянном разделении памяти и вычисления. Субъективная реальность демонстрирует принципиально иное устройство памяти, а соответственно, способов обработки информации (среди них есть и такие, которые можно назвать вычислениями), способов ассоциативной, контекстуально значимой выборки определенного контента из памяти и замены его другим (в активированных динамических контурах рефлексивного и арефлексивного, актуального и диспозиционального). Но эти специфические для нашей психики формы информационных процессов определяются главными системно-структурно-функциональными свойствами сознания (субъективной реальности) с учетом непрерывной связи с бессознательными процессами.

6. Охарактеризуем ключевые функциональные структуры субъективной реальности и присущие ей *операциональные способности*, которые могут иметь непосредственное отношение к моделированию когнитивных архитектур, соответствующих уровню AGI.

Элементарный анализ раскрывает фундаментальное структурно-операциональное свойство всякого явления субъективной реальности – его ***двумерность***, т.е. способность одновременного отображения в нем объекта и вместе с тем самого себя, т.е. ***единство иноотображения*** и ***самоотображения.*** Субъективная реальность в любых ее интервалах и проявлениях обладает в той или иной степени способностью самоотобра-

жения на уровне нашего Я. Эта *фундаментальная* способность отчетливо проявляется на уровне зрительного восприятия, как показано у Дж. Риццолатти и его сотрудников в исследовании зеркальных нейронов. Зеркальные нейроны одновременно осуществляют «кодирование сенсорной информации в моторных терминах», т.е. «создают единство зрительного восприятия и его вероятной моторной репрезентации. Это свидетельствует о том, что в то время как индивид наблюдает за выполнением действий другими, в его мозге выполняется потенциальный моторный акт, по всем своим характеристикам сходный с тем, который спонтанно активируется при подготовке и выполнении реального действия. Различие лишь в том, что в одном случае это действие остается потенциальным, а в другом переводится в последовательность конкретных движений» [Риццолатти, Синигалья 2012, 97].

Благодаря нашим зеркальным системам, мы способны незамедлительно распознавать определенные типы действий, производимых другими, достигать ***понимания*** другого. Это указывает на особенности организации зрительного образа (объединения в нем отображения предмета и вероятного способа действия с ним, а также санкционирующего акта понимания, «принятия» смысла этого действия). Единство трех операций способно обеспечить эффективное взаимодействие между людьми. Нельзя не учитывать, что на моторных репрезентациях наблюдаемых действий, создаваемых зеркальными нейронами, ***«основана способность обучения при помощи подражания»*** [Риццолатти, Синигалья 2012, 97]. Это обстоятельство может служить стимулом для поиска способов построения моделей саморасширения диапазона «навыков» интеллектуального робота, его способности *умножать свои «навыки»* в данном классе действий, способности столь существенной для построения AGI.

Единство иноотображения и самоотображения выражено в многообразных формах, от развитой рефлексии до элементарного ***чувства принадлежности мне переживаемого мной в данный момент состояния*** субъективной реальности. Зачастую чувство принадлежности находится вне фокуса осознания,

слабо выражено; требует для его актуализации внимания. Оно особенно остро переживается во время его нарушения, в патологических случаях *деперсонализации*, когда больному кажется, что переживаемые им ощущения, мысли «не его», и он впадает в паническое состояние. В более тяжелых патологических случаях (например, при параноидальном бреде) больной уже не воспринимает нарушений процесса самоотображения, «не замечает» своей деперсонализации.

Данные психопатологии и психоневрологии предоставляют исключительно ценный материал для исследования самоотображения, убедительно раскрывают *взаимозависимость иноотображения* и *самоотображения,* демонстрируют различные варианты и стадии его нарушения, особенно ярко в тех случаях, когда состояние деперсонализации вызывает явления *дереализации*, нарушения отображения внешних объектов и собственного тела (см., например: [Меграбян 1962; Рамачандран 2014; Литвак 2007; Назлоян 2002]).

7. Взаимозависимость иноотображения и самоотображения проявляется в функционировании памяти (в регистрах рефлексивного и арефлексивного, актуального и диспозиционального). Полнее всего она выражена на уровне многомерной динамической структуры нашего Я, т.е. в том, что можно назвать ***базисной динамической структурой субъективной реальности.*** Она представляет собой ***единство и переменное соотнесение противоположных модальностей «Я» и «не-Я»***. Это единство представлено в каждом наличном интервале субъективной реальности, оно формирует его ценностно-смысловой каркас и деятельно-волевые векторы. В динамическом биполярном контуре «Я» – «не-Я» совершается движение «содержания» (контента) субъективной реальности. Это «содержание» способно переходить из модальности «Я» в модальность «не-Я» и наоборот (например, когда «содержание», относящееся к модальности «Я», мои личностные свойства, мои оценки себя становятся для меня объектом внимания, анализа и оценки, а значит, выступают в данном интервале уже в модальности «не-Я» и т.п.). Соответственно, свойства другого человека или

определенной социальной общности могут «присваиваться» мной, переходя в модальность «Я» (*категории контентов Я*). Такого рода взаимопреобразования, перемена модальности переживаемого «содержания» – это *механизм эффективного отображения и освоения действительности*, в том числе субъективной реальности, ее саморегуляции, а вместе с тем механизм освоения социального опыта и осуществления творческой деятельности.

Взаимопереходы модальностей «Я» и «не-Я» постоянно сохраняют биполярную структуру субъективной реальности в любом ее интервале, не нарушая идентичности персонального Я (исключая психопатологию). *Каждая из модальностей определяется лишь через противопоставление другой и соотнесение с ней*. Поэтому в самом общем виде «Я» есть то, что противополагается «не-Я» и соотносится с ним; и наоборот, «не-Я» есть то, что противополагается «Я» и соотносится с ним. Первостепенный интерес для понимания операциональных структур системы «Я», «Самости» представляют исследования выдающегося нейропсихолога Антонио Дамасио, прежде всего его анализ автобиографической самости [Дамасио 2018].

Базисная динамическая структура субъективной реальности раскрывается конкретнее, когда в ходе феноменологического анализа выявляются основные виды противопоставления и соотнесения «Я» с «не-Я». Если взять за систему отсчета модальность «Я», то в первом приближении «Я» выступает по отношению к «не-Я» как отношение «Я»: 1) к внешним объектам, процессам (внешней предметности); 2) к собственному телу; 3) к самому себе; 4) к другому «Я» (другому человеку); 5) к «Мы» (той социальной общности, группе, с которой «Я» себя идентифицирует, к которой оно себя в том или ином отношении причисляет, ценности которой разделяет); 6) к «Они» (той общности, социальной группе, которой «Я» себя противопоставляет или, по крайней мере, от которой оно себя отделяет, ценности которой резко отрицает или равнодушно не принимает, не понимает); 7) к «Абсолютному» («Мир», «Бог», «Космос», «Природа» и т.п.).

Таков один из мыслимых способов выделения основных видов «содержания» (контента) «не-Я», а следовательно, самого

«Я», поскольку оно полагает и раскрывает себя посредством своего «не-Я». Иными словами, таковы главные смысловые (когнитивные и ценностные) измерения нашего «Я» (*категории контентов Я*). Чтобы полнее раскрыть одно из выделенных отношений «Я», нужно рассмотреть его не только само по себе, но и сквозь призму всех остальных. Нельзя основательно понять отношения «Я» к самому себе, если оставить в стороне отношение «Я» к предметному миру, к собственной телесности, к другому «Я», к «Мы», к «Они» и к «Абсолютному». В этой биполярной динамической контентной многомерности «Я» и полагает себя как свое «не-Я», выступающее в форме *«знания», «оценки»* и «*действия*». В ней непрерывно совершаются процессы самоотображения и самоорганизации структуры субъективной реальности, формируется и реализуется ее активность, ее ***деятельно-волевые функции***. Наиболее значимый вопрос о связи отображения с действием, как в форме самопознания и самопреобразования сознания, так и в форме познания и преобразования внешнего мира, заслуживает отдельного обсуждения в дальнейшем.

8. Приведенные результаты феноменологического анализа динамической структуры субъективной реальности для специалистов в области ИИ, особенно для тех, кто нацелен на разработку AGI, могут быть полезны во многих отношениях. Прежде всего, речь идет о ясном понимании той огромной дистанции, которая отделяет нынешний уровень ИИ от ЕИ, от конкретных специфических функциональных свойств естественного процесса мышления, к моделированию и реализации которых у нас пока нет технологического доступа. Но, занимая реалистическую позицию, мы сохраняем веру в наши *творческие* способности для того, чтобы добиваться шаг за шагом сближения с ЕИ, т.е. ***добиваться все более глубокого самопознания и самопреобразования, т.к. эти функции исходно присущи нашему ЕИ***.

На каждом этапе мы используем методологические установки, позволяющие корректно сокращать чрезвычайную многомерность субъективной реальности (в ее динамических

операциональных и творческих проявлениях), выбирать и осваивать некоторые ее специфические функциональные свойства, опираться на приемлемые аналогии в операциональных действиях мышления человека и интеллектуального робота. Лучше говорить в этом отношении не об ИИ в общем смысле, а об интеллектуальном роботе, поскольку он обладает определенной когнитивной архитектурой и, кроме того, совершает физически реальные действия во внешней среде, подлежащие строгой проверке на их эффективность.

Обратимся к рассмотренной выше двумерности явлений сознания: ***единство иноотображения и самоотображения – необходимое условие эффективности мыслительных и практических действий человека.*** Интеллектуальный робот обладает в известной мере аналогичными функциями. У него есть четко заданное иноотображение (и соответствующий набор внешних действий). Существует и нечто, подобное самоотображению, которое может быть заложено в надежности его программного обеспечения и технического устройства, в механизмах слежения за их исправностью. Но сразу при этом видим и качественное различие. У человека функция самоотображения обладает способностью *саморегуляции*, которая постоянно стимулируется, корректируется, совершенствуется процессами иноотображения, как и наоборот: стимулируется, корректируется, повышает результативность под воздействием процессов самоотображения, т.е. обладает способностью саморегуляции. В этом и заключается свойственное сознанию, структуре субъективной реальности единство функций иноотображения и самоотображения, их органическая взаимообусловленность, которая особенно ярко проявляется на уровне базовой динамической бимодальной структуры субъективной реальности – в единстве и переменном соотнесении модальностей «Я» и «не-Я».

Когда говорят об автономности роботов, то в полной мере осознают ее ограниченность: узость заданного диапазона восприятий и действий, неспособность к самостоятельному его расширению при существенном изменении внешних условий. Еще сильнее такая ограниченность связана с функцией

самоотображения. В программном обеспечении робота и в его техническом устройстве нет специальных регистров текущего непрерывного самоотображения его подсистем, процессов и соответственной самонастройки, подверженных непрестанной коррекции со стороны иноотображения и внешних действий, что, однако, наблюдается у человека. Функция иноотображения и то, что условно можно было бы отнести к функции самоотображения, фактически *разделены.*

Это указывает на проблемы, требующие разработки при создании интеллектуальных роботов нового поколения, т.е. достигающих уровня AGI. Полагаем, определенные ресурсы уже существуют, и наметились тенденции, связанные с критическим взглядом на ограниченность классической методологии А. Тьюринга признанием важной роли учета феноменов сознания в разработке ИИ.

В моделировании интеллектуальных роботов могут использоваться результаты исследования сознания, что ведет к расширению «мира» робота, его «техно-умвельта», тем самым и диапазона его успешных действий, согласно концепции А.Р. Ефимова, известного отечественного специалиста в области интеллектуальной робототехники [Efimov 2020]. В этом направлении, видимо, будет возможность использовать ресурсы, связанные с методами глубоких нервных сетей для создания программ указанной выше *двумерной функциональности* в поведении интеллектуального робота.

Разумеется, пунктирно намеченные возможности новых подходов нуждаются в дальнейшей конкретизации, критическом обсуждении. Но они свидетельствуют о широкой творческой перспективе применения феноменологических исследований операциональной специфики сознательной деятельности.

9. Дополним, что изложенные выше результаты феноменологических исследований субъективной реальности заставляют обратить внимание на чрезвычайную сложность «содержания» (множества различных *контентов*) нашей сознательной деятельности, взаимосвязь, взаимовлияние различных по своему виду (классу) контентов. Они приблизительно определены нами

в качестве категориальных ценностно-смысловых структур, отображающих области внешней предметности и процессуальности, телесности человека, собственного Я, понимания Я другого человека, социально и мировоззренчески обусловленных ценностно-смысловых и деятельно-волевых структур, рассмотренных нами ранее в статье.

Эти различные категориальные структуры ***диспозиционально*** укоренены в нашем сознании и на бессознательном уровне психики. Они обычно вовлечены в той или иной степени во все мыслительные процессы, задают широкий по их «содержанию» *спектр операций абстрагирования*, *анализа и синтеза.* Такие операции существенно различаются по когнитивным свойствам в зависимости от того или иного категориального контента. С одной стороны, операции абстрагирования, анализа и синтеза при исследовании предметной области внешнего мира, с другой – при исследовании нашего Я или социокультурных реалий.

Различие «онтологий» обусловливает соответственные особенности эпистемологических подходов. Тем не менее при решении сложных теоретических и особенно практических задач естественным интеллектом, как правило, используется широкий спектр этих операций. Многие из них протекают совместно, включают *интуитивный уровень*, связанный порой с новыми, находящимися в стадии формирования творческими результатами. Из многообразного спектра в системах ИИ используется лишь небольшая часть когнитивных операций, но в достаточной степени определенных. В этом состоит одно из главных отличий ЕИ от ИИ. Все изложенные обстоятельства должны служить поводом для размышлений при разработке AGI, поскольку в отношении него не требуется способность решения ***всего разнообразия*** человеческих задач (например, объяснения природы темной энергии и т.п.). Но важно уточнять те их классы, которые соответствуют его целям и возможностям на нынешнем этапе. Среди них особую актуальность в последнее время приобретают этические проблемы развития систем ИИ, которым следует посвятить основательное самостоятельное исследование.

10. В многомерных планах проблемы сознания целесообразно выделить еще один вопрос. Он касается не просто когнитивных операций (абстрагирования, обобщения и др.) в различных категориальных контентах, указанных выше, а учета множества *экзистенциально значимых* состояний сознательной деятельности, которые, хотя и включены в процесс познания, играют в нем существенную роль, но не могут быть четко выражены в принятых когнитивных терминах. Все они, однако, требуют учета и охвата для наличия целостного образа сознания, полноты анализа познавательного процесса и оценки его целей, т.к. в настоящее время, в условиях нарастания глобального кризиса нашей потребительской цивилизации, особое значение приобретают вопросы о подлинных смыслах личной и социальной деятельности, о противостоянии нарастающему абсурду в нынешних процессах социального развития.

Неслучайно известный феноменолог М. Мерло-Понти писал: «Жизнь сознания – познающая жизнь, жизнь желания, или жизнь перцептивная – скрепляется “интенциональной дугой”, которая проецирует вокруг нас наше прошлое, будущее, наше житейское окружение, нашу физическую, идеологическую и моральную ситуации или, точнее, делает так, что мы оказываемся вовлечены во все эти отношения. Эта интенциональная дуга и создает единство чувств, единство чувств и мышления, единство чувствительности и двигательной функции» [Мерло-Понти 1999, 182]. В данном случае понятие *интенциональной дуги* выполняет интегративную роль, т.е. роль средства охвата и концептуального учета диапазона «участников» познавательного процесса, от перцептивного уровня до идеологических, моральных и иных, экзистенциально значимых составляющих познавательного процесса, часть которых нередко остается в тени.

М. Мерло-Понти не назвал в приведенном перечне эстетическую составляющую, ***эстетическую интенцию***, которая в той или иной степени всегда присутствует в познавательных процессах (часто в неявной форме, но тем не менее дает о себе знать в виде неравнодушия познающего субъекта к красоте и безобразию, к тому, что называют возвышенным и низменным,

трагическим и комическим, гармоничным, изящным и т.п.). Все это приобретает высокую степень выражения и ценности в профессиональном художественном познании, в творчестве выдающихся представителей литературы и искусства. Если обратиться к интересующей нас проблеме сознания в широком смысле, то, полагаем, справедливо признать, что наиболее глубокие и разносторонние сведения о сознании даны в произведениях Пушкина, Толстого, Чехова, других писателей и поэтов.

Велико значение поэзии для эпистемологического анализа обыденного и научного мышления. Поэтические образы и метафоры – мощный творческий стимул развития интеллекта, т.к. разрушают устоявшиеся комфортные клише, решительно и убедительно пересекают границы привычных противоположных смыслов, обнаруживают ранее не известные степени свободы в нашем мышлении, создают в нем новые ценностно-смысловые образования и векторы активности, просветляют наш духовный мир.

Эстетическая интенция в познавательных процессах не менее значима, чем этическая, что находит отражение и в контексте ЕИ, и при обсуждении проблем разработки AGI. Следует иметь в виду не только когнитивный аспект, способ решения определенной задачи, но и оценку ее допустимости с позиций этических норм и принятых социальных критериев.

Особенно важная и трудная тема в проблематике развития ИИ – выяснение и использование для целей AGI операций бессознательно-психического уровня и допсихического, генетически заданного уровня, на котором осуществляется управление сложнейшими процессами в организме. Биологическая эволюция выработала многочисленные информационные компетенции эффективного управления сложными системами, которые унаследованы человеком, но остаются малоисследованными. Подтверждением служит следующий пример. В нашем теле присутствует 100 000 километров кровеносных русел, и эта колоссальная динамическая система эффективно управляется при постоянном изменении внешних физических условий.

Заключение

В завершение еще раз обратим внимание на ключевую роль образования в системе мер, необходимых для создания AGI, о чем говорил в обращении к читателям первой книги об AGI Герман Греф. Подготовка специалистов в области ИИ обнаруживает существенные недостатки. Авторы книги убеждены в следующем: «Современные образовательные программы готовят инженеров, умеющих работать с существующими технологиями, а не творчески мыслящих исследователей, способных совершать новые прорывы» [Сильный ИИ… 2021, 55]. В этом проявляется слабость, формализм, интеллектуальное убожество (в ряде отношений) современного образования. Чтобы воспитывать творческую устремленность, оно должно делать акцент на фундаментальных теоретических вопросах развития ИИ, формировать у молодого инженера широкий духовный кругозор, интерес к философским и социальным проблемам познания и самопознания, к поэзии, гениальной художественной литературе как источникам и стимулам творческого мышления.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Дамасио 2018 – *Дамасио А.* Так начинается «Я». Мозг и возникновение сознания. – М.: Карьера-Пресс, 2018.

Дубровский 1971 – *Дубровский Д.И.* Психические явления и мозг. Философский анализ проблемы в связи актуальными задачами нейрофизиологии, психологии и кибернетики. – М.: Наука, 1971.

Дубровский 2002 – *Дубровский Д.И.* Проблема идеального. Субъективная реальность. –М.: Канон+, 2002.

Дубровский 2015 – *Дубровский Д.И.* Проблема «Сознание и мозг». Теоретическое решение. – М.: Канон+, 2015.

Дубровский 2019 – *Дубровский Д.И.* Субъективная реальность и мозг: к вопросу о полувековом опыте разработки проблемы сознания в аналитической философии // *Дубровский Д.И.* Проблема сознания. Теория и критика альтернативных концепций. – М.: Ленанд, 2019. С. 140–183.

Дубровский 2020 – *Дубровский Д.И.* Субъективная реальность // Философская электронная энциклопедия. – М.: Институт философии РАН, 2020. – URL: https://iphlib.ru/library/collection/elphilenc/document/dubrovskii_subektivnaya_realnost.

Кузнецов 2016 – *Кузнецов О.П.* Избранные труды. Автоматы, языки и искусственный интеллект. – М.: Наука, 2016.

Литвак 2007 – *Литвак Л.М.* «Жизнь после смерти»: предсмертные переживания и природа психоза. Опыт самонаблюдения и психоневрологического исследования. – М.: Канон+, 2007.

Меграбян 1962 – *Меграбян А.А.* Деперсонализация. – Ереван: Армгосиздат, 1962.

Мерло-Понти 1999 – *Мерло-Понти М.* Феноменология восприятия. – СПб.: Ювента; Наука, 1999.

Назлоян 2002 – *Назлоян Г.М.* Концептуальная психотерапия. Портретный метод. – М.: ПЕР СЭ, 2002.

Рамачандран 2014 – *Рамачандран В.С.* Мозг рассказывает. Что делает нас людьми. – М.: Карьера Пресс, 2014.

Риццолатти, Синигалья 2012 – *Риццолатти Дж., Синигалья К.* Зеркала в мозге. О механизмах совместного действия и сопереживания. – М.: Языки славянских культур, 2012.

Сильный ИИ… 2021 – Сильный искусственный интеллект. На подступах к сверхразуму / А.А. Ведяхин и др.; науч. ред. А.С. Потапов. – М.: Интеллектуальная литература, 2021.

Финн 2011 – *Финн В.К.* Искусственный интеллект: методология, применения, философия. – М.: URSS, Красанд, 2011.

Финн 2016 – *Финн В.К.* О классе ДСМ-рассуждений, использующих изоморфизм правил индуктивного вывода // Искусственный интеллект и принятие решений. 2016. № 3. С. 48–61.

Dubrovsky 2019a – *Dubrovsky D.I.* "The Hard Problem of Consciousness." Theoretical Solution of Its Main Questions // AIMS Neuroscience. 2019. Vol. 6. No. 2. P. 85–103.

Dubrovsky 2019b – *Dubrovsky D.I.* The Problem of Free Will and Modern Neuroscience // Neuroscience and Behavioral Physiology. 2019. Vol. 49. No. 5. P. 629–639.

Efimov 2020 – *Efimov A.* Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence // Artificial General Intelligence. 13th International Conference, AGI 2020 (Lecture Notes in Computer Science. Vol. 12177) / ed. by B. Goertzel, A. Panov, A. Potapov, R. Yampolskiy. – Cham: Springer, 2020. P. 83–94.

Koch, Massimini, Boly, Tononi 2016 – *Koch C., Massimini M., Boly M., Tononi G.* Neural Correlates of Consciousness: Progress and Problems // Nature Rewies Neuroscience. 2016. Vol. 17. No. 5. P. 307–321.

REFERENCES

Damasio A. (2010) *Self Comes to Mind: Constructing the Conscious Brain, Pantheon.* New York: Pantheon Books (Russian translation: Moscow: Karyera-Press, 2018).

Dubrovsky D.I. (1971) *Mental Phenomena and the Brain. Philosophical Analysis of the Problem in Connection with the Urgent Tasks of Neurophysiology, Psychology, and Cybernetics.* Moscow: Nauka (in Russian).

Dubrovsky D.I. (2002) *The Problem of the Ideal. Subjective Reality.* Moscow: Kanon+ (in Russian).

Dubrovsky D.I. (2015) *The Problem "Consciousness and the Brain." Theoretical Solution.* Moscow: Kanon+ (in Russian).

Dubrovsky D.I. (2019a) "The Hard Problem of Consciousness". Theoretical Solution of Its Main Questions. *AIMS Neuroscience.* Vol. 6, no. 2, pp. 85–103.

Dubrovsky D.I. (2019b) The Problem of Free Will and Modern Neuroscience. *Neuroscience and Behavioral Physiology.* Vol. 49, no. 5, pp. 629–639.

Dubrovsky D.I. (2019c) Subjective Reality and the Brain: On the Issue of the Semi-Centennial Experience in the Development of the Problem of Consciousness in Analytical Philosophy. In: Dubrovsky D.I. *The Problem of Consciousness. Theory and Criticism of Alternative Concepts* (pp. 140–183). Moscow: Lenand (in Russian).

Dubrovsky D.I. (2020) Subjective Reality. In: *Philosophical Electronic Encyclopedia.* Moscow: RAS Institute of Philosophy. Retrieved from https://iphlib.ru/library/collection/elphilenc/document/dubrovskii_subektivnaya_realnost

Efimov A. (2020) Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence. In: Goertzel B., Panov A., Potapov A., & Yampolskiy R. (Eds) *Artificial General Intelligence. 13th International Conference, AGI 2020* (Lecture Notes in Computer Science. Vol. 12177) (pp. 83–94). Cham: Springer.

Finn V.K. (2011) *Artificial Intelligence: Methodology, Applications, Philosophy.* Moscow: URSS, Krasand (in Russian).

Finn V.K. (2016) On the Class of JSM-Reasoning Using Isomorphism of Inductive Inference Rules. *Artificial Intelligence and Decision Making.* No. 3, pp. 48–61 (in Russian).

Koch C., Massimini M., Boly M., & Tononi G. (2016) Neural Correlates of Consciousness: Progress and Problems. *Nature Rewies Neuroscience.* Vol. 17, no. 5, pp. 307–321.

Kuznetsov O.P. (2016) *Selected Works. Automata, Languages and Artificial Intelligence.* Moscow: Nauka (in Russian).

Litvak L.M. (2007) *"Life after Death": Near-Death Experiences and the Nature of Psychosis. The Experience of Self-Observation and Neuropsychiatric Research.* Moscow: Kanon+ (in Russian).

Megrabyan A.A. (1962) *Depersonalization.* Yerevan: Armgosizdat (in Russian).

Merleau-Ponty M. (1945) *La Phénoménologie de la perception.* Paris: Gallimard (Russitan translation: Saint Petersburg: Yuventa; Nauka, 1999).

Nazloyan G.M. (2002) *Conceptual Psychotherapy. Portrait Method.* Moscow: PER SE (in Russian).

Ramachandran V.S. (2011) *The Tell-Tale Brain: A Neuroscientist's Quest for What Makes Us Human.* New York: W.W. Norton & Company (Russian translation: Moscow: Karyera-Press, 2014).

Rizzolatti G. & Sinigaglia C. (2008) *Mirrors in the Brain: How Our Minds Share Actions and Emotions.* Oxford: Oxford University Press (Russian translation: Moscow: LRC Publishing House, 2012).

Vedyakhin A.A. et al. (2021) *Strong Artificial Intelligence. On the Approaches to the Supermind.* Moscow: Intellektual'naya literatura (in Russian).

DOI: 10.30727/0235-1188-2021-64-1-45-70
Оригинальная исследовательская статья
Original research paper

Новое пифагорейство, сложность, искусственный интеллект

Т.С. Ахромеева
Институт прикладной математики имени М.В. Келдыша РАН, Москва, Россия

Г.Г. Малинецкий
Институт прикладной математики имени М.В. Келдыша РАН, Москва, Россия

Аннотация

Лейбниц определил математику как «науку о возможных мирах». Но какие миры полагались возможными мыслителями разных эпох? Именно этот вопрос и рассматривается в настоящей работе. По мысли Эйнштейна, развитие науки требует «внешнего оправдания» (наблюдений и экспериментов, которые следует объяснить или предсказать) и внутреннего совершенства (следования внутренней логике этой дисциплины). У математики нет «внешнего оправдания» и эксперимент не может опровергнуть математическую теорию. И в этом смысле математика ближе к творчеству, к искусству, чем к естественной науке. Ее связь с культурой оказывается более сложной и опосредованной. Для объяснения этой связи Д. Белл в рамках своей теории общественного развития выдвигает «осевой принцип», в соответствии с которым в качестве основной характеристики общества рассматривается роль науки в его развитии. С этой точки зрения выделяется традиционное, индустриальное и постиндустриальное общества. Каждой из этих фаз соответствуют свои идеалы, нормы и типы математического творчества. В традиционном обществе, следуя пифагорейской традиции, математика ориентирована на поиск гармонии в природе, на выявление единства на основе универсальных соотношений, определяемых числовыми характеристиками изучаемых сущностей. С приближением к индустриальной эпохе все большее значение приобретает конструктивистский, «проектный» подход. И единство возникает на более высоком «метауровне». Предтечей этого направления предстает является Декарт, поставивший вопрос о поиске единого, универсального метода решения всех математических задач. В работе прослежено изменение постановок ряда «инвариантных», «вечных» математических проблем, а также эволюция понятия «сложности» в исторической

ретроспективе. Основное внимание уделено постиндустриальной фазе развития цивилизации и «компьютерной математике», которая стала основой формирования виртуальной реальности, во многом изменившей само направление прогресса. В результате этого «экстравертная ориентация» человечества, курс на новые горизонты сменила «интровертная», которая во главу угла ставит задачи, связанные с комфортом, удобством, потреблением. Произошедшая «смена вех» прослежена на основе сравнения «больших проектов», связанных с математикой, выдвигавшихся в 1960-е годы, и тех, которые рассматриваются сейчас как приоритетные. По сути дела, мы столкнулись с «кризисом ожиданий». Выход из этого кризиса видится в возрождении на новом уровне «пифагорейской традиции». Но если в традиционном обществе целью развития математики в союзе с другими науками и искусствами было выявление гармонии в мире природы, то в постиндустриальной фазе приоритеты иные. Они связаны с компьютерным моделированием, пониманием и выявлением основ гармонии в мире людей.

Ключевые слова: философия математики, междисциплинарные подходы, гуманитарно-технологическая революция, постиндустриальное общество, цифровизация реальности, стратегический прогноз.

Ахромеева Татьяна Сергеевна – кандидат физико-математических наук, научный сотрудник Института прикладной математики имени М.В. Келдыша РАН.
akhromeyeva@gmail.com
http://orcid.org/0000-0002-8340-5796

Малинецкий Георгий Геннадьевич – доктор физико-математических наук, профессор, заведующий отделом моделирования нелинейных процессов Института прикладной математики имени М.В. Келдыша РАН.
gmalin@keldysh.ru
http://orcid.org/0000-0001-6041-1926

Для цитирования: *Ахромеева Т.С., Малинецкий Г.Г.* Новое пифагорейство, сложность, искусственный интеллект // Философские науки. 2021. Т. 64. № 1. С. 45–70. DOI: 10.30727/0235-1188-2021-64-1-45-70

New Pythagoreanism, Complexity, Artificial Intelligence

T.S. Akhromeyeva
Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia

G.G. Malinetskiy
Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia

Abstract

Leibniz defined mathematics as the "science of possible worlds." But what worlds were assumed as the possible ones by thinkers of different eras? It is this question that is considered in this article. According to Einstein, development of science requires "external justification" (observations and experiments that should be explained or predicted) and internal perfection (following the internal logic of this discipline). Mathematics has no "external justification" and experiment cannot refute mathematical theory. In this sense, mathematics is closer to creativity, to art than to natural science. Its connection with culture turns out to be more complex and mediated. To explain this connection, Daniel Bell, within the framework of his theory of social development, puts forward the "axial principle," according to which the role of science is considered as the main characteristic of society. From this point of view, we single out traditional, industrial and post-industrial societies. Each of these phases has its own ideals, norms, and types of mathematical creativity. In traditional society, following the Pythagorean tradition, mathematics is focused on finding harmony in nature, on identifying unity on the basis of universal relationships determined by the numerical characteristics of the objects under study. As the industrial era approaches, the constructivist, or "project-based," approach becomes increasingly important. And unity emerges at a higher "meta-level." The forerunner of this direction is Descartes, who raised the question of finding a single, universal method for solving all mathematical problems. The work traces the change in the formulation of a number of "invariant," "eternal" mathematical problems as well as the evolution of the concept of "complexity" in the historical retrospective. Close attention is paid to the post-industrial phase of civilizational development and to "computer mathematics," which has become the basis for formation of virtual reality, which in many respects changes the very direction of progress. As a result of this, the "extraverted orientation" of humanity, the course toward new horizons was replaced by the "introvert" one, which prioritizes tasks associated with comfort, convenience, and consumption. The "change of milestones" that has taken place is traced in a comparison of "big projects" related to mathematics that were put forward in the 1960s, and those that are considered as priorities at present. In fact, we are faced with a "crisis of expectations." We see a way out of this crisis in a revival of the "Pythagorean tradition," at a new level. But if in a traditional society the goal of the development of mathematics in alliance with other sciences and arts was to reveal harmony in the natural world, then at the post-industrial phase, the priorities are different. They

are associated with computer modeling, understanding, and identifying the foundations of harmony in the human world.
Keywords: philosophy of mathematics, interdisciplinary approach, humanitarian and technological revolution, post-industrial society, digitalization of reality, strategic forecasting.

Tatiana S. Akhromeyeva – Ph.D. in Physics and Mathematics, Research Fellow, Keldysh Institute of Applied Mathematics, Russian Academy of Sciences.
akhromeyeva@gmail.com
http://orcid.org/0000-0002-8340-5796

Georgiy G. Malinetskiy – D.Sc. in Physics and Mathematics, Professor, Head of the Department of the Modeling of Non-linear Processes, Keldysh Institute of Applied Mathematics, Russian Academy of Sciences.
gmalin@keldysh.ru
http://orcid.org/0000-0001-6041-1926

For citation: Akhromeyeva T.S. & Malinetskiy G.G. (2021) New Pythagoreanism, Complexity, Artificial Intelligence. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 45–70.
DOI: 10.30727/0235-1188-2021-64-1-45-70

Постановка проблемы

> Когда человек не знает, к какой пристани он держит путь, для него ни один ветер не будет попутным.
> *Сенека Младший.* LXXI письмо Луцилию

Уже в XX веке наука стала большой отраслью экономики и важным стратегическим ресурсом, определяющим путь страны в будущее и ее национальную безопасность. Профессия исследователя стала массовой. При этом фронт научных работ многократно расширился. Если «научные сверхдержавы» – СССР и США – вели исследования по всему фронту, активно развивая все области, то сейчас такой возможности нет. Приходится выбирать, ориентируясь не только на научные интересы исследователя, но и на те задачи, которые решает общество, и на перспективу. И чем дальше, тем важнее становится такой «стратегический подход» к науке, взгляд на нее «с птичьего полета». В 1960-х годах господствовало ощущение, что «нам нет преград ни в море, ни на суше», или, говоря словами Гильберта, произнесенными на Математи-

ческом конгрессе в Париже в 1900 году: «Мы хотим знать. Мы будем знать». Сейчас взгляд существенно изменился – все чаще ученые сталкиваются с проблемами, лежащими за пределами их возможностей или с принципиальными ограничениями.

В этих заметках мы посмотрим «с птичьего полета» на тенденции развития математики в контексте тех задач, которые в разные эпохи решало человечество, и на связанные с ними технологии. Возможно, это окажется полезным при выборе приоритетов развития цифровой реальности.

Математическая парадигма и традиционное общество

> Знание некоторых принципов легко возмещает незнание некоторых фактов.
>
> *А. Гельвеций*

> 1 000 000 000 000 066 600 000 000 000 001 – простое число, называемое также числом Бельфегора.

В 1950-х годах американский социолог Даниел Белл выдвинул так называемый «осевой принцип». В соответствии с ним историю человечества можно проектировать на разные оси и получать ее различное разбиение на сменявшие друг друга эпохи. В качестве такой оси он выбрал роль и место знания в развитии общества. В этой проекции выделяется традиционное общество, развивающееся почти до XX века, индустриальное общество, определявшее прошлое столетие, и постиндустриальное общество, переход к которому происходит в настоящее время. Это приводит к глубоким сдвигам в сознании, в мировосприятии, в картине мира и в целях, которые ставит перед собой наука: «На протяжении большей части человеческой истории *реальностью была природа:* и в поэзии, и в воображении люди пытались соотнести свое “я” с окружающим миром. Затем *реальностью стала техника*, инструменты и предметы, сделанные человеком, однако получившие независимое существование вне его “я”, в овеществленном мире. В настоящее время *реальность является, в первую очередь, социальным миром* – не природным, не вещественным, а исключительно человеческим – воспринимаемым через отражение своего “я” в других людях… Поэтому неизбежно, что постиндустриальное общество ведет к появлению нового уто-

пизма, как инженерного, так и психологического. Человек может быть переделан или освобожден, его поведение – запрограммировано, а сознание – изменено. Ограничители прошлого исчезли вместе с концом эры природы и вещей» [Белл 1999, 663].

Почти весь пройденный человечеством исторический путь был связан с традиционным обществом. Людей при этом волновал достаток, защита и понимание реальности, которое с ними связано.

Принципиальный шаг был сделан выдающимся философом и математиком Пифагором (ок. 580 – ок. 500 до н.э.). Он провозгласил: «В основе всего – число». По его мысли, в основе окружающего мира, а также таких нематериальных понятий, как музыка и справедливость, лежат числовые пропорции. В реальности есть глубокая внутренняя связь между различными явлениями, гармония, описываемая математическими соотношениями. Ее необходимо постичь, чтобы овладеть всеми тайнами мироздания. Доказанная в его научной школе теорема Пифагора наглядно показывала, что между некоторыми числами и фигурами действительно есть удивительные связи. Историки математики насчитывают более 250 доказательств этой прекрасной теоремы.

Евклид решал конкретную задачу о числе существующих правильных многогранников в дополнение к известным (тетраэдр, куб, октаэдр, додекаэдр, икосаэдр), создал аксиоматический метод и выдвинул идею, определившую на 20 с лишним веков вектор развития математических наук и всего естествознания. Это идея о том, что очень многие свойства нашего мира, а в идеале всю реальность, можно осмыслить и описать, опираясь на немногие принципы (аксиомы). Многие века многочисленные успехи в развитии этого подхода связывались с промыслом божьим.

Например, идеалом современной теоретической физики является создание картины реальности, опираясь только на мировые константы: заряд электрона e, скорость света c, постоянную Планка h, гравитационную постоянную G и, возможно, на одну или несколько величин, связанных с сильными и слабыми взаимодействиями. Но может быть, удастся описать наш мир, пользуясь еще меньшим числом постоянных?

Влияние «математических мотивов» на мировоззрение античного мира было огромно. На воротах Платоновской Академии, просуществовавшей более 900 лет, было начертано: «Да не войдет сюда тот, кто не знает геометрии».

Аристотель полагал, что наука, в отличие от искусства, не имеет практического значения. Глубокие внутренние связи и стоящая за ними гармония важны сами по себе. Одна из замечательных теорем из «Начал» Евклида утверждает, что число простых чисел в натуральном ряду бесконечно…

Наш современник, специалист по теории чисел Ю.И. Манин высказывает парадоксальную мысль о возможной связи этого раздела математики – теории чисел – и математических констант – e (основание натуральных логарифмов), π и других с физическими постоянными и теми инструментами, которые используются в теоретической физике [Манин 2010]. По сути дела, это развитие той же самой пифагорейской традиции на новом уровне.

Понимание гармонии в науке существенно отличается от ее трактовки в философии, религии или искусстве. Это понимание уточнил принцип, выдвинутый в средние века У. Оккамом, утверждающий: «Не следует создавать сущности сверх необходимости», или в формулировке Эйнштейна: «Все должно быть сделано как можно проще, но не проще этого».

Столетием взлета науки и технологий стал XVII век. Именно здесь возникла прочная связь между математикой – наукой о числах и фигурах – и естествознанием. Галилео Галилей считал, что «Книга Природы написана языком математики», и выдвинул свой императив: «Измерить все, что является измеримым, и сделать таковым то, что измеримым не является». И действительно, именно совершенствование инструментов для измерения и осмысления полученных результатов стало магистральным пунктом развития науки.

Например, Кеплеру, чтобы установить, что планеты вращаются по эллипсам, а не по окружностям, нужно было знать их орбиты с точностью в несколько процентов. Лапласу – астроному наполеоновской эпохи, – чтобы учитывать в небесной механике влияние одних планет на другие, нужны были уже в 1 000 раз более точные данные.

Уже в Античности легенды рассказывали о практических приложениях идей мудрецов. Первому вошедшему в историю философу – Фалесу Милетскому – приписывали удивительно точный прогноз богатого урожая олив (что недоступно нынешним метеорологам), позволивший скупить маслобойни, а затем получить большой денежный выигрыш. В историю вошел не только архимедов винт, но и легенда о том, как Архимед с помощью зеркал сжег вражеский флот…

Мир сложен и справиться с этим в течение многих веков помогали гуманитарные технологии (принятие разумных законов, выбор способа правления, адекватного состоянию общества и т.д.) и изобретательность. Первые механические часы, каравеллы, которые пересекали Атлантику, и величественные строительные сооружения появились не благодаря расчетам науки, а благодаря интуиции, фантазии и накопленному за много поколений опыту.

С XVII века благодаря математическому моделированию, основы которого заложил Ньютон, ситуация меняется. Наука все чаще оказывается готова прийти на помощь практике. И уже Эйлер рассчитывает несущую балку парусных кораблей, что позволяет сделать их более легкими, маневренными и дает им решающее преимущество в бою.

В XVII веке была высказана идея, становящаяся пророческой. Выдающийся немецкий математик, физик, юрист, языковед, философ и дипломат Г.В. Лейбниц (1646–1716) изобрел калькулятор, который мог умножать и делить числа. (Справедливости ради следует сказать, что выдающийся философ, публицист и математик Блез Паскаль (1623–1662) в том же столетии изобрел счетный аппарат. Видимо, эта идея витала в воздухе.) И, по мысли Лейбница, «считающие машины» в будущем станут настолько точными, объективными и мудрыми, что им можно будет поручить судопроизводство. Машины будут судить людей. Именно сейчас это пророчество воплощается в системах социального рейтингования, в которых искусственный интеллект на основе имеющихся у него данных (покупки, разговоры, перемещения) выставляет оценку каждому человеку. На основе этих оценок, в конечном итоге, власти определяют социальный статус и доступные человеку, обладающему им, возможности. Кроме того, Лейбниц определил математику как «науку о возможных мирах». Подчеркнем – не о нашем конкретном, а обо всех, которые мы можем представить и исследовать.

Тогда же были намечены две главные линии, по которым развивалось естествознание. Первая – это анализ (дословно: дробление, расчленение), поиск элементарных сущностей (в физике – элементарные частицы, в биологии – клетки, в экономике – товар и т.д.). Из свойств элементарных простейших частей исследователи надеялись вывести свойства целого, и во многих случаях это удавалось. Второй путь – в космос, ко все большим масштабам, от астрономии к космологии.

Исследования в обоих направлениях продолжают развиваться, за очередные достижения и в физике элементарных частиц, и в космологии продолжают присуждаться Нобелевские премии, но, судя по всему, исследовательский потенциал этих направлений исчерпывается.

В математику XX века было введено очень важное понятие *некорректных задач*. Это проблемы, решение которых либо не существует, либо они неустойчивы по отношению к вариации начальных данных. Немного изменив их, мы получим совсем другой ответ.

Дело в том, что фундаментальная физика сегодня имеет дело с процессами и пробует предложить их теории, которые разворачивались на временных масштабах, сравнимых с «планковским» временем – 10^{-44} с и планковским масштабом – 10^{-35} м. Одна из общепринятых теорий космологии, объясняющая большой взрыв, который привел к формированию нашей вселенной, имеет дело с процессами, разворачивающимися на временах порядка 10^{-36} с. Мегамасштабы, которыми оперирует космология, простираются до 13,5 млрд световых лет.

Чем ближе к этим границам, тем меньше становится измерений, экспериментов, наблюдений, позволяющих выбрать между конкурирующими теориями. Чем ближе к границам, тем сложнее, дороже становятся инструменты, которые позволяют проводить такие измерения, тем меньше их становится. Есть шанс, что Большой адронный коллайдер станет *последним* инструментом такого масштаба, в необходимости которого физики смогли убедить общество.

Результаты развиваемых сейчас теорий также парадоксальны. Исследователи, развивающие петлевую теорию квантовых гравитаций, приходят к выводу, что на рассматриваемых ими микромасштабах времени нет. Та сущность, которую мы называем временем, в рамках этой теории связана с несовершенством нашего восприятия и тем, что мы оказались в том месте космоса, где энтропия возрастает [Ровелли 2020].

Джордано Бруно отстаивал утверждение о множественности миров. Теория инфляции вводит в рассмотрение мультиверс – множество вселенных, возникающих в момент большого взрыва, которые принципиально недоступны нашему наблюдению: «Вечная инфляция предсказывает, что наша Вселенная (физическая область пространства, откуда свет успевает дойти до нас за

14 млрд лет, с момента нашего Большого взрыва) – это лишь одна из бесконечного числа вселенных мультиверса I уровня, в котором происходит все, что может произойти» [Тегмарк 2019, 219].

Это удивительно похоже на ту картину, которую представляли себе некоторые философы: «Демокрит полагал, что бесконечность состоит из одинаковых миров, где люди неизменно исполняют одни и те же судьбы. Паскаль (на которого, кроме всего прочего, могли повлиять слова Анаксагора, что каждая вещь заключает в себе все вещи) располагает эти сходные миры один за другим, так что в пространстве не остается атома, который не включал бы вселенную, ни вселенной, которая не была бы также атомом» [Борхес 1992, 49].

Круг замкнулся. Мыслители Античности полагали, что философы могут разрабатывать метафизику – науку о наивысших, последних началах бытия, доступных лишь умозрительно. Ряд физических теорий оперирует именно такими, недоступными для экспериментального исследования сущностями.

Разумеется, и в математике, имеющей отношение к природе, и в физике, на долю и нашего, и следующих поколений остается много открытых проблем. Однако границы уже очерчены, и математика новой эпохи будет связана с другими задачами.

Индустриальная эпоха и «считающие машины»

Мы хотели, чтобы появились летающие машины.
А получили твиты на 140 символов.
П. Тиль (основатель системы
онлайн-платежей PayPal)

Прошедшее столетие без преувеличения можно назвать веком техники и технологий. Именно они обеспечили небывалый рост благосостояния и увеличение продолжительности жизни для миллиардов людей. Антибиотики и родовспоможение на более высоком, чем раньше, уровне уменьшили смертность. Прямой синтез аммиака из воздуха, открытый Ф. Хабером в 1914 году, позволил производить азотные удобрения и во многом снял проблему голода. XX век вошел в историю как время тотального использования машин в самых разных областях.

С точки зрения математики, он прошел удивительно быстро. «Миры», которые строили и исследовали математики, «отставали» от стремительно менявшейся реальности. Достаточно посмотреть

на 23 проблемы, поставленные Д. Гильбертом в 1900 году, которым математики уделили огромное внимание в XX веке, чтобы убедиться, насколько далеки они от задач, которые решало индустриальное общество. Императивами эпохи стали массовость, стандартизация, взаимозаменяемость. Ее математическим символом – *алгоритм* – фиксированная последовательность действий, приводящая к решению поставленной задачи.

По-видимому, важнейшим достижением математики XX века и одним из главных технологических прорывов стало создание компьютеров. Они многократно расширили пределы наших возможностей.

Следуя логике Лейбница, выдающийся британский математик и криптограф Алан Тьюринг в 1930-х годах предложил универсальную конструкцию, позволяющую формализовать любые вычислительные процедуры, названную впоследствии машиной Тьюринга. И вновь можно проследить удивительное сходство с проблемой, которую в XVII веке ставил Декарт в работе «О методе». В этой работе он стремился выяснить, существует ли общий метод решения *всех* математических задач. Декарт полагал, что такой метод существует.

Машина Тьюринга представляет собой головку, которая может находиться в нескольких состояниях, и бесконечную ленту, на которой написаны нули, единицы и оператор STOP. Работа машины заключается в том, что головка считывает символ перед ней и в зависимости от него и своего внутреннего состояния передвигает ленту вправо или влево, сохраняя или меняя символ, находящийся перед ней. Тезис Тьюринга состоит в том, что все возможные алгоритмы вычислений могут осуществляться с помощью этой машины.

Фундаментальный результат Тьюринга состоит в том, что Декарт был не прав, – существуют задачи, которые не могут быть решены при помощи такой машины [Пенроуз 2011].

Эту работу, ответившую на вопрос великого философа, можно назвать пророческой. Все современные компьютеры (за исключением нейрокомпьютеров) представляют собой несовершенную реализацию машины Тьюринга. Она была создана «на кончике пера», задолго до того, как появилась механическая или электрическая база, позволившая ее воплотить в том или ином виде.

Машина, алгоритм, технология стали важнейшими образами науки и культуры индустриальной эпохи. Можно вспомнить

тезис выдающегося архитектора Ле Корбюзье, считавшего, что «Дом – это машина для жилья», памятник III Коммунистическому интернационалу В.Е. Татлина, и, конечно, конвейер. В центре внимания этой эпохи оказались компьютеры. Если пифагорейцы искали гармонию в природе, определяемую количественными соотношениями, придавали числам мистическое значение, то в компьютерный век их место заняли алгоритмы, позволяющие с помощью вычислительных машин решать задачи, недоступные человеку. Машины сыграли огромную роль в проектировании машин же, в решении инженерных задач – в расчетах траекторий ракет и космических аппаратов, в атомном проекте, в создании микросхем, где на кристалле следует разместить миллиарды элементов.

Компьютеры многократно расширили пределы наших возможностей и изменили представления о сложности. Во времена Пифагора и через много веков после него сложными математическими проблемами считались те, которые требовали много времени и усилий либо для того, чтобы найти их решение, либо доказать, что его не существует.

Один из популярных в наше время подходов трактует эту категорию иначе. Пусть входные данные, необходимые для решения задачи, представляют собой массив длины N (символов, нулей и единиц, цифр, букв какого-либо алфавита). Если число операций, которые нужны, чтобы получить ответ Q, растут в зависимости от N степенным образом $Q \sim N^{\alpha}$, то такие задачи называют простыми или полиномиальными. Все остальные проблемы называют *сложными*.

Пример сложной проблемы дает задача коммивояжера. Она состоит в том, чтобы, двигаясь с постоянной скоростью, объехать 70 городов, расположенных на плоскости. Простейший способ решать задачу – воспользоваться перебором. Первый город мы можем выбрать N способами, для выбора второго есть N–1 вариант, третьего N–2 и т.д. Таким образом общее число вариантов $N\cdot(N-1)\cdot\ldots 2\cdot 1 = N!$. Эта величина растет с увеличением N быстрее, чем $(N/e)^N$ ($e = 2{,}718\ldots$). Для $N = 70$ $N! \approx 10^{100}$. Это число, названное физиком-теоретиком Ф. Дайсоном гуголом, очень велико. По его мысли, такие числа не должны возникать в физических исследованиях. По оценкам космологов, в нашей вселенной около 10^{82} атомов.

Существуют иные способы решения этой проблемы. Алгоритм Хелда–Карпа позволяет найти кратчайший маршрут за $2^N\cdot N^2$

операций [Стюарт 2015]. Как видим, задача и при таком подходе остается сложной. Естественно, возникает общий вопрос, напоминающий декартовский, а является ли эта (и все другие) задача на самом деле сложной? Или для ее решения можно придумать полиномиальный алгоритм? Эта проблема является одной из самых сложных в современной математике. Пока строго доказать, что действительно существуют сложные задачи, не удается [Стюарт 2015].

Здесь стоит упомянуть еще одну гипотезу, связанную с вычислениями и выдвинутую британским математиком и физиком С. Вольфрамом. Пояснить ее можно так. Когда дети учатся считать, им объясняют сложение с помощью палочек. Здесь можно действовать по определению и пересчитывать палочки. Ответ получается верный, но дело это долгое. Можно действовать быстрее, складывая числа «в столбик». Такие задачи, в которых можно действовать быстрее, чтобы узнать, «что будет в конце», будем называть *вычислительно приводимыми*. Те же, где такого быстрого способа нет, и к ответу можно прийти только, действуя по определению, выполняя вычисления «в лоб», шаг за шагом (это называется имитационным моделированием), – *вычислительно неприводимыми*. Вольфрам предположил, что наука может добиться успеха только в вычислительно приводимых задачах или в тех, которые могут быть к ним сведены. Например, в физике аналитические результаты получаются, когда объектов немного (уже задача трех тел требует компьютерных вычислений), либо когда они все одинаковые (типичный подход статистической физики и квантовой механики). В других случаях (биологическая эволюция, динамика экономических систем, моделирование элементов сознания и многих других) короткого пути нет, и нам, чтобы узнать ответ, приходится на компьютерах шаг за шагом просчитывать математическую модель этой системы [Wolfram 1983].

Индустриальная эпоха изменила приоритеты математики. Если с XVII по XX век в центре внимания были непрерывные системы, рассматривая которые можно иметь дело с действительными числами, бесконечно малыми величинами, то в конце «фаворитами» стали разделы, имеющие дело с дискретными, «цифровыми» величинами, целыми числами.

Казалось бы, индустриальная эпоха, как в свое время традиционное общество, должно было породить свою оригинальную философию, позволяющую осмыслить перспективы и импера-

тивы развития техники и технологий, их влияния на общество и человека. К сожалению, этого не произошло – эпоха кончилась слишком быстро. К исключениям можно отнести выдающегося философа, футуролога XX века Станислава Лема. Он так комментирует этот феномен «несостоявшейся философии»: «Я стал философом в то время, когда в королевстве философии уже невозможно строить большие системы, потому что это королевство распалось из-за вторжений науки, так что философ в результате не может быть независимым создателем картины мира. Поэтому он должен соглашаться на сосуществование с наукой или, напротив, ему остается отступление в изоляцию, как уже произошло с феноменологией или языковой философией… Научная фантастика означала для меня научную строгость и в то же время еще и привилегию творческой свободы, которую предоставляет искусство. То есть, несмотря ни на что, все-таки означала философию, пусть и скрытую переводом в литературу» [Лем 2018, 279].

Императивом философского осмысления индустриальной эпохи являлась *кибернетика*. Этот междисциплинарный подход задумывался его создателями как общая теория управления и связи в машине, обществе, человеке. Во многом этот подход представлял собой взгляд инженера и математика на все эти сущности. В книге «Сумма технологии» – своеобразной «библии индустриального общества» – Лем раскрыл «человеческое измерение» технологического процесса и представил предельные технологические возможности человечества.

Почему век «технологического оптимизма», экспансии человечества в космос, в океан, микромир оказался недолог? Не обсуждая эту важную и интересную проблему в целом, поясним ее на примере авиации. На начальном этапе развитие шло по очевидной схеме – быстрее, выше, сильнее. Но уже сверхзвуковые лайнеры «Конкорд» и Ту-144 столкнулись с серьезными экономическими, экологическими и инфраструктурными ограничениями. Поэтому мы летаем на тех же высотах и с теми же скоростями, что и 60 лет назад.

С боевой авиацией ситуация иная – после периода быстрого совершенствования начался переход к беспилотникам. Человек оказался самым медленным, уязвимым и ненадежным звеном в боевых машинах. Переход к беспилотникам Лем предсказывал задолго до того, как он произошел. Но он предвидел и следующие шаги. На смену беспилотникам приходят роботы, способные к

коллективным действиям, стаи «кремниевой саранчи», делающие большинство видов современных вооружений бесполезными. Затем открытое противостояние оказывается слишком опасным, наступает эпоха гибридных войн, в которых противостоящие стороны делают акцент на бактериях и вирусах…

Вехой конца индустриальной эпохи стал фантастический отказ от пилотируемой космонавтики. Последний полет на Луну имел место почти полвека назад. Вместо экспансии человека в космос, в глубины океанов, вглубь вещества был сделан поворот к компьютерному миру и виртуальной реальности. Экстравертную ориентацию цивилизации сменила интровертная…

Математика постиндустриального общества и искусственный интеллект

> Компьютеры – прекрасная вещь, они позволяют автоматизировать многие задачи – практически все, что не требует понимания. Само понимание находится вне их. Компьютеры не понимают.
>
> *Р. Пенроуз*

Казалось бы, при переходе к постиндустриальному обществу, к миру людей математика должна измениться. Но среди решавшихся «великих задач» или «задач тысячелетия» в этой области нет *ни одной,* связанной с анализом поведения людей или обществом [Стюарт 2015]. Видимо, для постановки фундаментальных задач, отражающих изменившуюся реальность, нужны гораздо бо́льшие характерные времена.

Вместе с тем, с человеческим фактором специалисты по моделированию сталкиваются на каждом шагу. Простейший пример – «многокритериальная оптимизация» или попросту задача выбора. Ее суть показывает известная в прошлом юмореска, в которой один герой предлагает купить другому крупных раков, но по пять рублей, или совсем мелких, но по три. Другой герой никак не может решить, что же для него важнее – цена раков или их размер. Это именно та задача, с которой регулярно сталкиваются конструкторы и руководители. Обычно компьютерные модели позволяют сказать, к чему приведет то или иное решение, но выбор остается за ними людьми. Поэтому надо разбираться, как же человек принимает решения.

Например, есть интересная теория, утверждающая, что мы делаем это при помощи простейшего вероятностного алгоритма, опирающегося на теорему Байеса [Дехан 2016].

Пусть нас интересует, какова вероятность события B при условии, что событие A уже наступило $P(B|A)$.

Пусть вероятность события A – $P(A)$, события B – $P(B)$, а вероятность события A при условии, что произошло событие B – $P(A|B)$. Тогда по теореме Байеса

$$P(B|A) = P(A|B) \cdot P(A) / P(B). \quad (1)$$

Мир дает нам хаотическую последовательность сигналов на их основе, и на основании того, чему мы научились, мы оцениваем вероятности $P(A)$, $P(B)$ и $P(A|B)$. На этой основе оцениваем $P(A|B)$. *Когда эта вероятность становится велика*, то мы заключаем, что на основании сигнала A можно сделать вывод, что произошло B. Осталось лишь разобраться, как наш мозг формирует $P(A)$, $P(B)$ и $P(A|B)$…

Другой пример, показывающий, что люди могут вести себя парадоксальным антиинтуитивным образом, дают эпидемии. Классическая теория дает для зависимости числа заразившихся от времени так называемое логистическое уравнение $dn/dt = \alpha n(N-n)$, где $n(t)$ – число людей заразившихся к моменту t, а $n(0) = n_0$, где α – коэффициент, показывающий, насколько быстро будет расти число заболевших, n_0 – число заразившихся в начале эпидемии, N – предельное число заразившихся в конце. Зная решение этого уравнения, данные по зависимости $n(t)$ вначале, определяют α и N.

Но коэффициент α зависит от того, объявлен ли карантин и насколько население склонно его соблюдать. На этом основании и делаются многие прогнозы. Но общество, например, в Москве, отреагировало парадоксально, – когда $n(t)$ было мало, оно соблюдало карантин и следовало рекомендациям врачей. Когда $n(t)$ выросло, оно стало соблюдать их хуже, коэффициент α увеличился, а с ним и N… Если бы у нас были хорошие социологические модели и понимание такого поведения, то прогнозы были бы точнее, но пока этого нет…

Стоит обратить внимание на подход к анализу «человекомерных систем», называемому *агентным моделированием*. В таких моделях людям противопоставляются компьютерные «агенты», для которых определяются правила взаимодействия, а затем производится компьютерный расчет и выясняется, к чему приводят сделанные предположения. Например, такие модели позволяют

описать распределение богатства в обществе. Эти результаты находятся в разительном противоречии с «очевидными» догмами либеральной экономики.

Генеральный секретарь ООН Антониу Гуттериш в начале 2020 года указал рост имущественного неравенства как одну из четырех главных угроз, с которыми столкнулся современный мир. И действительно, сегодня 8 богатейших людей мира владеют таким же количеством активов, как и 3,5 млрд беднейших жителей планеты [Стиглиц 2016].

Многоагентные модели показывают, что для объяснения этого явления вполне достаточно учесть несовершенство операций обмена, когда покупатель немного переплачивает за товар или соглашается на меньшую сумму [Богосян 2020].

Суть дела поясняет один из парадоксов теории вероятностей, связанный с игрой в казино. Предлагается следующая игра. Вначале на кон ставится, например, 100 руб. и если выпадет орел, то на кону оказывается 120 руб., если решка – 83. Добавлять или забирать деньги со стола в течение всего вечера нельзя. Стоит ли вступать в такую игру?

На первый взгляд, конечно стоит. Если монета «честная», то орел и решка выпадают с вероятностью ½, и ожидаемый выигрыш $20 \cdot ½ - 17 \cdot ½ = 1{,}5$.

Но если игра продолжается, то картина меняется. Если монету бросили 10 раз и в пяти из них выпал орел, а в пяти – решка, то на кону окажется $(1{,}2)^5 \cdot (0{,}83)^5 = 98{,}02$. То есть такая «честная» игра ведет к проигрышу. И чем дальше она будет продолжаться, тем больше будет проигрыш. Агентное моделирование позволяет учесть множество осложняющих факторов и показать, при каких условиях возникает олигархия. И это очень хорошо согласуется с результатами по целому ряду рассмотренных стран.

В свое время выдающийся математик, философ, мыслитель Н.Н. Моисеев писал, что управление обществом на основе математических моделей, планирование экономики на современном компьютерном уровне и использование разработок, уже выполненных в Академии наук, может дать экономический эффект, сравнимый с результатами работы всей страны в течение пятилетки [Моисеев 1979]. К сожалению, этим надеждам не дано было осуществиться. Модели социально-экономических и экологических процессов, проекты управления народным хозяйством на основе компьютерного анализа и прогноза оказались не

востребованы советской системой государственного управления. Реформы последних 30 лет исходили из ненужности подобных подходов.

Более того, лауреат Нобелевской премии по экономике Роберта Солоу в 1987 году заметил, что массовое использование компьютеров в экономике не привело к повышению производительности труда ни в одной отрасли промышленности, кроме производства компьютеров. Многие эксперты полагают, что мировая экономика сейчас находится в «зоне Солоу».

Однако компьютер сыграл очень важную социальную роль. С одной стороны, он позволил «сжигать» свободное время миллиардов человек, снабжая их ненужной информацией, развлекая компьютерными играми и общением в социальных сетях. Оказалось, что очень много людей в мире готовы проводить значительную часть своей жизни в призрачном, виртуальном мире, не имеющем отношения к их насущным проблемам.

С другой стороны, интернет и современные информационно-телекоммуникационные технологии сделали неравенство очевидным и тем самым обострили противоречия между богатым Севером и бедным Югом, между рядом стран и регионов [Контуры цифровой реальности… 2018].

Мир начала постиндустриальной эпохи, тесно связанного с цифровизацией, стал гораздо сложнее индустриальной реальности. Во-первых, он стал намного быстрее, многие глобальные социально-экономические процессы могут привести к кардинальным переменам в течение месяцев, а то и недель. Во-вторых, увеличилась связность мира. Увеличилось число активных игроков на «большой шахматной доске» и их влияние друг на друга, даже если они находятся в разных концах планеты. В-третьих, намного большую роль в нем играют рефлексивные процессы. Существенно не только событие, но и его восприятие, а также восприятие восприятия. Именно поэтому огромную власть приобретают СМИ, Интернет, социальные сети. Именно здесь, а не в экономике, науке или планировании цифровизация играет ключевую роль, однако серьезный системный или математический анализ этих процессов только начат [Проектирование будущего… 2018].

В настоящее время большие надежды связываются с искусственным интеллектом. По сравнению с естественным интеллектом определить ИИ гораздо проще. В его основе лежат алгоритмы распознавания образов или классификации, которые

разрабатываются уже более полувека. В них по информации на входе компьютер должен опознать тип сигнала и отнести к одному из классов. Например, по записи звука «разобрать» сказанное слово по буквам, или по набору результатов анализов отнести данного пациента к больным или здоровым и т.д. Главная идея ИИ связана с так называемыми нейронными сетями и их глубоким обучением. Другими словами, в таких компьютерных системах нам не следует писать программу, определяющую, какой сигнал можно отнести в какой класс. В них мы задаем «правило обучения», позволяющее таким системам «самим» учиться на примерах [Малинецкий, Потапов, Подлазов 2016].

По сути дела, это то самое «байесовское обучение», связанное с многократным использованием формулы (1). Огромное быстродействие и память позволяют иметь дело с более сложными сигналами, анализ которых доступен людям. С большим количеством образов, с более быстрыми и точными действиями, поскольку соответствующие вероятности машина может оценить точнее, чем человек. Хотя теория нейронных сетей разрабатывалась давно, но технологии здесь обогнали науку, а сам ИИ превращается в своеобразную отрасль промышленности, которая в недалекой перспективе освободит людей от рутинной умственной работы.

В настоящее время системы с ИИ обыгрывают людей в шахматы и го, дают справки, используются в банках и кадровых агентствах, водят автомобили.

Вместе с тем способность решать задачи в очень небольшой мере соответствует тому, что мы называем сознанием. Философ Джон Серль доказывает, что никакой цифровой компьютер не может обладать состояниями сознания (если понимать под состоянием сознания, к примеру, то, что происходит у вас в голове, когда я говорю: «Представьте себе красную розу» – и вы ее себе представляете): сознание нельзя построить из компьютерной программы Цифровой компьютер способен лишь выполнить тривиальные арифметические и логические операции… Представьте, что вы сортируете все более и более толстую колоду. Можете представить, что в какой-то момент, когда стопка станет достаточно толстой, появится некое новое сознание? «Да ни за что!» – пишет известный специалист по информатике Д. Гелернтер [Гелернтер 2016, 363].

Тем не менее мы имеем дело с серьезной проблемой. Она неоднократно рассматривалась и в мифах, и в антиутопиях. Ученик

волшебника вызвал силы, не представляя себе риски, которые с ними связаны. И очень важно, используя ИИ, не попасть в такую ситуацию. Починка машины (10 тыс. деталей) или обслуживание самолета (около 100 тыс.) является непростой задачей, хотя имеется детальная документация и понятно, как устроены эти машины. С компьютерными программами в десятки мегабайт ситуация хуже. Исследования показывают, что человек в среднем делает, по крайней мере, одну ошибку на 1 000 команд. И это приводит, например, к наличию десятков тысяч «уязвимостей» в широко распространенных операционных системах, к феномену хакерства и периодическим эпидемиям компьютерных вирусов. Тем не менее здесь разработчики понимают, что должны делать и как работает тот или иной фрагмент программы.

В случае систем ИИ ситуация значительно хуже. Мы не знаем, какие решающие правила и стратегии вырабатывает нейронная сеть на основе предъявленного ей набора примеров. Поэтому вполне возможно, что тестирование систем с ИИ станет одной из массовых профессий уже в недалеком будущем. Тем не менее во многих важных задачах, связанных с высоким риском, с принятием стратегических решений, принимаемых в оборонном комплексе «в режиме реального времени», от систем с ИИ следует отказываться. Как заметил в свое время Станислав Лем, «неслыханно быстрые машины ошибаются неслыханно быстро». При этом договоры об отказе от таких систем (так же, как в свое время об отказе от химического и биологического оружия) следует начать как можно быстрее.

Системы с ИИ во многих случаях подобны оракулу или волшебной палочке. Не вполне понятен результат их использования. Машина и человек «думают» по-разному. Поэтому приходится либо доверять тому, что не понимаем, а это весьма опасно, либо воспринимать результат работы ИИ как «добрый совет» и самому выстраивать логическое обоснование принимаемых решений.

И это только небольшая часть социальных проблем, порожденных этой технологией, воплотившей мечту Лейбница. Поэтому игра в го, в шахматы, рутинная умственная работа, где цена ошибки не очень велика, вполне подходит для этой «математической игрушки». Очевидно, компьютерные системы следует ориентировать на возможности, потребности и ограничения людей, а не наоборот, как бы ни были заманчивы человеко-машинные утопии.

Заключение. Протокол о намерениях

> Люди грядущего поколения будут знать многое, неизвестное нам, и многое останется неизвестным для тех, кто будет жить, когда изгладится всякая память о нас. Мир не стоит ломаного гроша, если в нем когда-нибудь не останется ничего непонятного.
>
> *Сенека Младший.* О природе (VII, 30, 5)

Конец индустриальной и начало постиндустриальной эпохи можно назвать временем инноваций. Постоянные перемены стали нормой. Общество освоило «технологии создания технологий». Новшества прогнозируют, планируют, разрабатывают, а потом внедряют. Но сами эти прогнозы и ожидания «специального заказа» очень многое говорят об обществе, о желаемом будущем.

И с этой точки зрения интересно взглянуть на планируемые технологии в части, непосредственно связанной с математическими разработками. Это позволит прочертить нынешний вектор развития цивилизации.

В 2019 году Фраунгоферовское общество (общество содействия прикладным исследованиям), крупнейшее научное объединение Европейского союза, сделало технологический прогноз.

Его целью было выявить, какие направления будут привлекать наибольший объем инвестиций в 2030 году. Из 300 тем были выбраны 30, самых перспективных по мнению экспертов. Отобранные технологии разбили на 4 группы по степени готовности. Обратим внимание на несколько из них, особенно тесно связанных с математикой.

Первая группа включает технологии, которые разрабатываются уже сейчас. Самый высокий рейтинг эксперты присвоили глубокому обучению нейронных сетей. В таких сетях есть несколько слоев, каждый из которых соответствует своему уровню абстракции. Предполагают, что они смогут гораздо эффективнее обрабатывать разнородные данные, включающие статистические и динамические изображения, звуки, тексты. В этой связи можно вспомнить глубокое замечание Станислава Лема о том, что на самом деле человечеству нужен не «искусственный интеллект», а «искусственный инстинкт».

Во вторую группу попала постквантовая криптография. Здесь мы имеем извечную борьбу щита и меча. Власти, элиты,

спецслужбы ряда стран стремятся создать «прозрачный мир», в котором все действия любого человека фиксируются и навсегда записываются. Люди и сообщества, ценящие личное пространство и определенную свободу, стремятся защититься от тотального контроля. По прогнозам многих ученых, квантовые компьютеры позволят вскрывать все шифры, созданные по современным стандартам. Поэтому защита конфиденциальной информации требует нового уровня технологий – постквантовых. Они также опираются на фундаментальные законы квантовой механики и воплощают известный принцип – «лечить подобное подобным».

К технологиям второй группы, которые планируют разработать к 2030-м годам относят также кибернетические рассуждающие системы. Они должны реализовать известную мечту программистов – искать уязвимости в созданных ими программах и самостоятельно устранять их. Это должно повысить безопасность информационной инфраструктуры будущего.

В третью группу вошло то, что будет реализовано существенно позже и сегодня воспринимается как научная фантастика – квантовые коммуникации и квантовые компьютеры, искусственный мозг, нейронные чипы. Искусственный мозг основан на идее, что когда нейронов в нейросети станет достаточно много (в мозге человека их около 86 млрд) и у них будет много связей (в мозге до 1500), то количество перейдет в качество, и произойдет следующая компьютерная революция. Нейроморфные чипы должны уметь работать с информацией в режиме реального времени – например, быть идеальными синхронными переводчиками или существенно расширять возможности наших органов чувств.

В четвертой группе оказались технологии, ориентированные на решение социальных проблем и сохранение окружающей среды. Они лишь косвенно используют результаты математических наук [Комаров 2019].

Не менее интересно и то, что не вошло в список приоритетов. В них НЕТ термоядерного синтеза, освоения космоса, создания городов в океане; нет технологий предотвращения войн и борьбы с терроризмом, борьбы с инфекциями и эпидемиями. В них нет прямой задачи продления активной здоровой жизни человека, управления погодой, альтернативной энергетики и новых средств транспорта.

Фактически все, о чем мечтали фантасты середины XX века, выпало из списка приоритетов. Без большого преувеличения можно

сказать, что компьютерная реальность и искусственный интеллект оттеснили на периферию сознания научного и инженерного сообщества наиболее важные проблемы человека и общества.

Фантасты 1960-х годов мечтали о том, чтобы раздвинуть горизонты человечества, грезили о пути к звездам, а нынешние эксперты делают ставку на безопасность и комфорт…

Ситуация удивительно напоминает ту, которая была описана Станиславом Лемом в книге «Возвращение со звезд» (1961), которая сейчас воспринимается как пророческая. В ней группа космонавтов возвращается из дальнего космоса через век с лишним после отлета. За этот век Земля неузнаваемо изменилась. Перед отлетом императивом была космическая экспансия, исследование дальнего космоса. По возвращении они застали цивилизацию комфорта и развлечений, которой не нужны те знания, добыванию которых они посвятили жизни [Лем 2019].

Отказ человечества от амбициозных проектов, который показывают нам последние десятилетия, вложение усилий не в обычную, а в виртуальную реальность, «обмен развития на комфорт» уже обошелся очень дорого.

Это кризис в науке, деградация системы образования, торможение технологического развития. И как следствие всего этого – нарастание социально-экономических и экологических проблем.

Вероятно, шанс сменить нынешний регресс на будущий прогресс дает изменение вектора интереса человечества, своеобразное *новое пифагорейство, возвращение к звездам*.

Пифагор надеялся найти гармонию в природе, осмыслить единство всего сущего, опираясь на математические образы [Клайн 1984]. И уже в XX веке Эйнштейн писал: «Весь предшествующий опыт убеждает нас в том, что природа представляет собой реализацию простейших математических мыслимых элементов, я убежден, что посредством чисто математических конструкций мы можем найти те понятия и закономерные связи между ними, которые дадут нам ключ к пониманию явлений природы… Конечно, опыт остается единственным критерием пригодности математических конструкций физики. Но настоящее творческое начало присуще именно математике. Поэтому я считаю оправданной в известном смысле веру древних в то, что чистое мышление в состоянии постигнуть реальность» [Эйнштейн 1967, 484].

Гуманитарно-технологическая революция заставляет нас не ограничиваться природой, а расширить поле поисков, включив

в него человека, сознание, общество, многие проблемы, считавшиеся ранее лежащими в области гуманитарных и философских наук.

Возможен ли такой поворот? Ведь предыдущие 25 веков математика шла рука об руку с философией природы. Впрочем, историк и методолог математики Макс Клайн извлек из пройденного пути еще один урок: «Математическая мысль без устали бьется о каменный берег, препятствующий ее проникновению на новые территории. Но даже гранитные утесы не выдерживают ее могучего натиска, не ослабевающего на протяжении столетий, и рушатся, открывая перед математикой новые просторы» [Клайн 1988, 236].

Думается, что уже в ближайшие годы мы убедимся, что оптимизм этого прогноза оправдан.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Белл 1999 – *Белл Д.* Грядущее постиндустриальное общество: Опыт социального прогнозирования. – М.: Academia, 1999.

Богосян 2020 – *Богосян Б.* Неизбежное казино // В мире науки. 2020. № 1/2. С. 130–139.

Борхес 1992 – *Борхес Х.Л.* Письмена Бога. – М.: Республика, 1992.

Гелернтер 2016 – *Гелернтер Д.* Счетные бесконечные множества и состояния сознания // Теории всего на свете / под. ред. Д. Брокмана. – М.: Лаборатория знаний, 2016 С. 362–363.

Дехан 2016 – *Дехан С.* Универсальный алгоритм принятия человеческих решений // Теории всего на свете / под ред. Д. Брокмана. – М.: Лаборатория знаний, 2016. С. 369–372.

Клайн 1988 – *Клайн М.* Математика. Поиск истины. – М.: Мир, 1988.

Комаров 2019 – *Комаров С.М.* Будущее, что стоит у двери // Химия и жизнь. 2019. № 12. С. 2–5.

Контуры цифровой реальности… 2018 – Контуры цифровой реальности. Гуманитарно-технологическая революция и выбор будущего / ред. В.В. Иванов, Г.Г. Малинецкий, С.Н. Сиренко. – М.: Ленанд, 2018.

Лем 2018 – *Лем С.* Литература и философия, или «Я хотел спасти мир» // Четвертые Лемовские чтения / отв. ред. А.Ю. Нестеров. – Самара: Самар. гуманит. акад., 2018. – С. 279–287.

Лем 2019 – *Лем С.* Возвращение со звезд. – М.: АСТ, 2019.

Малинецкий, Потапов, Подлазов 2016 – *Малинецкий Г.Г., Потапов А.Б., Подлазов А.В.* Нелинейная динамика: Подходы, результаты, надежды. – М.: URSS, 2016.

Манин 2010 – *Манин Ю.И.* Математика как метафора / 2-е изд. доп. – М.: МЦНМО, 2010.

Моисеев 1979 – *Моисеев Н.Н.* Математика ставит эксперимент. – М.: Наука, 1979.

Пенроуз 2011 –– *Пенроуз Р.* Новый ум короля: О компьютерах, мышлении и законах физики / 4-е изд. – М.: URSS, 2011.

Проектирование будущего… 2018 – Проектирование будущего. Проблемы цифровой реальности. Вып. 1 (8–9 февраля 2018 г., г. Москва) / под ред. Г.Г. Малинецкого. – М.: ИПМ им. М.В. Келдыша, 2018.

Ровелли 2020 – *Ровелли К.* Срок времени. – М.: АСТ, Corpus, 2020.

Стиглиц 2016 – *Стиглиц Дж.* Великое разделение: Неравенство в обществе, или что делать оставшимся 99% населения. – М.: Эксмо, 2016.

Стюарт 2015 – *Стюарт И.* Величайшие математические задачи / пер. с англ. Н. Лисовой. – М.: Альпина нон-фикшн, 2015.

Тегмарк 2019 – *Тегмарк М.* Жизнь 3.0: Быть человеком в эпоху искусственного интеллекта. – М.: Corpus, 2019.

Эйнштейн 1967 – *Эйнштейн А.* Собр. соч. Т. 4. – М.: Наука, 1967.

Wolfram 1983 – *Wolfram S.* Statistical Mechanics of Cellular Automata // Reviews of Modern Physics. 1983. Vol. 55. No 3. P. 601–644.

REFERENCES

Bell D. (1973) *The Coming of Post-Industrial Society: A Venture in Social Forecasting.* New York: Basic Books (Russian translation: Moscow: Academia, 1999).

Boghosian, B. M. (2020). The Inescapable Casino. *The Best Writing on Mathematics.* Vol. 18, pp. 15–28 (Russian translation in: *V mire nauki.* 2020. No. 1/2, pp. 130-139.)

Borges J.L. (1992) *The Writing of the God.* Moscow: Respublika (Russian translation).

Gelernter G. (2013) Denumerable Infinities and Mental States. In: Brockman J. (Ed.) *This Explains Everything: Deep, Beautiful, and Elegant Theories of How the World Works* (pp. 365–366). New York: HarperCollins Publishers (Russian translation in: Moscow: Laboratoriya znaniy, 2016).

Dehaene S. The Universal Algorithm for Human Decision Making. In: Brockman J. (Ed.) *This Explains Everything: Deep, Beautiful, and Elegant Theories of How the World Works* (pp. 372–375). New York: HarperCollins Publishers, 2013 (Russian translation in: Moscow: Laboratoriya znaniy, 2016).

Einstein A. (1967) *Collected Works* (Vol. 4). Moscow: Nauka (Russian translation)

Kline M. (1988) *Mathematics and the Search for Knowledge.* Oxford, UK: Oxford University Press (Russian translation: Moscow: Mir, 1988).

Komarov S.M. (2019) The Future That Stands at the Door. *Khimiya i zhizn'.* No. 12, pp. 2–5 (in Russian).

Ivanov V.V., Malinetskiy G.G., & Sirenko S.N. (2018) *Contours of Digital Reality. Humanitarian and Technological Revolution and Choice of the Future.* Moscow: Lenand (in Russian).

Lem S. (2018) Literature and Philosophy, or "How I Wanted to Save the Worlds." In: Nesterov A.Y. (Ed.) *The Fourth Lem Readings.* Samara: Samara Humanitarian Academy (Russian translation).

Lem S. (2019) *Return from the Stars.* Moscow: AST (Russian translation).

Malinetskiy G.G. (Ed.) (2018) *Designing the Future. Problems of Digital Reality. Issue 1 (February 8–9, 2018, Moscow).* Moscow: RAS Keldysh Institute of Applied Mathematics (in Russian).

Malinetskiy G.G., Potapov A.B., & Podlazov A.V. (Eds.) (2016) *Nonlinear Dynamics: Approaches, Results, Hopes.* Moscow: URSS (in Russian).

Manin Yu.I. (2010) *Mathematics as a Metaphor* (2nd ed.). Moscow: MCCME (in Russian).

Moiseev N.N. (1979) *Mathematics Sets up an Experiment.* Moscow: Nauka (in Russian).

Penrose R. (1989) *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics.* Oxford: Oxford University Press (Russian translation: Moscow: URSS, 2011).

Rovelli C. (2017) *L'ordine del tempo.* Milan: Adelphi (Russian translation: Moscow: AST, Corpus, 2020).

Stewart I. (2013) *The Great Mathematical Problems: Marvels and Mysteries of Mathematics.* London: Profile Books (Russian translation: Moscow: Alpina non-fiction, 2015).

Stiglitz J.E. (2015) *The Great Divide: Unequal Societies and What We Can Do about Them.* New York: W.W.Norton & Company (Russian translation: Moscow: Exmo, 2016).

Tegmark M. (2017) *Life 3.0: Being Human in the Age of Artificial Intelligence.* London: Allen Lane (Russian translation: Moscow: Corpus, 2019).

Wolfram S. (1983) Statistical Mechanics of Cellular Automata. *Reviews of Modern Physics.* Vol. 55, no 3, pp. 601−644.

DOI: 10.30727/0235-1188-2021-64-1-71-87
Оригинальная исследовательская статья
Original research paper

Нейрофилософия, философия нейронаук и философия искусственного интеллекта: проблема различения

Е.А. Безлепкин
Институт философии и права Сибирского отделения РАН, Новосибирск, Россия

А.С. Зайкова
Институт философии и права Сибирского отделения РАН, Новосибирск, Россия

Аннотация

Под нейрофилософией понимают разные направления философии, в частности философию нейронаук, философию искусственного интеллекта или элиминативный материализм. Чрезмерная нагруженность термина связана с еще не завершившимся процессом осмысления предметной области этой дисциплины. К примеру, в одном из первых определений нейрофилософии П.С. Черчленд говорилось о редукции психологии к нейронаукам. С современных позиций представление о нейрофилософии как о попытке оправдать элиминативный материализм устарело и не соответствует действительности. В статье проведится анализ терминов «философия нейронаук», «нейрофилософия» и «философия искусственного интеллекта», а также предлагается вариант их разделения. Общность и различия показаны на примере теории сознания Дж. Эдельмана и концепции коннекционизма для слабого искусственного интеллекта. Делается вывод, что от интегрального использования понятия «нейрофилософия» целесообразно отказаться. Под термином «нейрофилософия» следует понимать направление в философии начала XXI века, применяющее нейронаучные концепции для решения традиционных философских проблем, а философия нейронаук может быть рассмотрена в первую очередь как раздел философии науки, который формулирует и решает проблемы и частных нейронаук, и нейронаучного направления в целом. «Философия искусственного интеллекта» – направление в философии, отвечающее на вопрос о том, что такое небиологический интеллект и возможен ли он. Иными словами, это – философско-методологическая база для изучения небиологического интеллекта. Авторы приходят к выводу, что в становлении нейронаук и их научно-философского базиса мы пока находимся на первом методологическом этапе анализа и диффе-

ренциации гипотез. Философия нейронаук как база существующих нейронаучных теорий возникнет через определенное время, и именно в соответствующий период будет необходим данный термин.

Ключевые слова: искусственный интеллект, нейронаучная философия, когнитивные науки, нейробиология, П.С. Черчленд, Дж. Эдельман, коннекционизм, нейросети.

Безлепкин Евгений Алексеевич – кандидат философских наук, научный сотрудник Института философии и права Сибирского отделения РАН,
evgeny-bezlepkin@mail.ru
https://orcid.org/0000-0001-9020-5445

Зайкова Алина Сергеевна – младший научный сотрудник Института философии и права Сибирского отделения РАН.
zaykova.a.s@gmail.com
https://orcid.org/0000-0003-3300-0130

Для цитирования: *Безлепкин Е.А., Зайкова А.С.* Нейрофилософия, философия нейронаук и философия искусственного интеллекта: проблема различения // Философские науки. 2021 Т. 64 № 1. С. 71–87.
DOI: 10.30727/0235-1188-2021-64-1-71-87

Neurophilosophy, Philosophy of Neuroscience, and Philosophy of Artificial Intelligence: The Problem of Distinguishing

E.A. Bezlepkin
Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science, Novosibirsk, Russia

A.S. Zaykova
Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science, Novosibirsk, Russia

Abstract

Neurophilosophy is understood as different areas of philosophy, for example, the philosophy of neuroscience, the philosophy of artificial intelligence, or eliminative materialism. This excessive interpretation of the term is due to the fact that the understanding of the subject area of this discipline is still incomplete. For example, one of the earliest definitions of neurophilosophy given by P.S. Churchland stated reduction of psychology to neurosciences. In modern views, the idea of neurophilosophy as an attempt to

justify eliminative materialism is outdated and does not correspond to reality. The article analyzes the terms "philosophy of neuroscience," "neurophilosophy," and "philosophy of artificial intelligence" and also offers a variant of their differentiation. The authors focus on the common and different features, using the example of G.M. Edelman's theory of consciousness and the concept of connectionism for weak artificial intelligence. It is concluded that integral use of the term "neurophilosophy" should be abandoned. As a result, the term "neurophilosophy" should be understood as a direction in philosophy of the early 21st century, applying neuroscientific concepts to solve traditional philosophical problems, while the philosophy of specific neurosciences can be considered primarily as a field in the philosophy of science that formulates and solves problems of specific neurosciences as well as of the entire neuroscientific direction. The philosophy of artificial intelligence is an area in philosophy that answers the question of what non-biological intelligence is and what makes it possible; in other words, it is a philosophical and methodological basis for the study of non-biological intelligence. In the formation of neurosciences and their scientific and philosophical basis, we are still at the first methodological stage of the analysis and differentiation of hypotheses. After some time, there will emerge a philosophy of neuroscience, as the basis of all existing neuroscientific theories, and then this term will acquire greater significance.

Keywords: artificial intelligence, neuroscientific philosophy, cognitive science, P.S. Churchland, G.M. Edelman, connectionism, neural networks.

Evgeny A. Bezlepkin – Ph.D. in Philosophy, Research Fellow, Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science.
evgeny-bezlepkin@mail.ru
https://orcid.org/0000-0001-9020-5445

Alina S. Zaykova – Junior Research Fellow, Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science.
zaykova.a.s@gmail.com
https://orcid.org/0000-0003-3300-0130

For citation: Bezlepkin E.A. & Zaykova A.S. (2021) Neurophilosophy, Philosophy of Neuroscience, and Philosophy of Artificial Intelligence: The Problem of Distinguishing. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 71–87.
DOI: 10.30727/0235-1188-2021-64-1-71-87

Введение

Философы сегодня по-прежнему спорят о том, что такое философия. В приведенном утверждении присутствует, конечно, доля

лукавства, поскольку определяющие критерии философии известны: предметная область достаточно четко определена, проблемы (как классические, так и современные) поставлены, стратегии их решения, по крайней мере, для философии частных наук, более или менее известны и алгоритмы их применения прописаны. Не хватает только четкого и емкого определения. Не хватает потому, что философия – необычайно разнородный внутри себя предмет, охватывающий области от истории философии до логики и философий конкретных наук.

Следовательно, любые попытки определения термина «философия» приводили и будут приводить к возникновению множества непреодолимых дискуссий. Подобное произошло и с термином «нейрофилософия». Вопрос об определении того, чем является нейрофилософия, можно соотнести с вопросом о том, что такое философия. Отсутствие определенности в постановке предметной области, проблем и методов их решения делает вопрос о том, что такое нейрофилософия, философской проблемой. В настоящей статье предлагается разграничение трех терминов: первый – нейрофилософия (нейронаучная философия), второй – философия нейронаук (т.е. философия частных наук, к которым применима приставка «нейро-») и третий – философия искусственного интеллекта.

Начнем с определения нейронауки. В самом общем смысле под нейронауками понимают междисциплинарную область знаний, которая изучает нейронные процессы. По сути, нейронауки начали отделяться от нейробиологии подобно тому, как в свое время физика и другие науки о природе начали выделяться из философии. Интересный формальный критерий для определения нейронаук предложил А.Ю. Алексеев: «Науке приписывается префикс “нейро-”, если в ее инструментарии преобладают коннекционистские методы исследования» [Алексеев, Кузнецов, Савельев, Янковская 2015, 55]. В данном случае термин «философия нейронаук» должен включать в себя всю философско-методологическую базу указанных наук. Итак, под это определение подходит и часть философии искусственного интеллекта. Действительно, направление, изучающее возможность создания искусственного интеллекта, включает в себя коннекционистские и символьные методы исследования. Если искусственный интеллект рассматривать как нейронауку, то философия искусственного интеллекта не может не являться частью философии нейронаук.

Ревонсуо в книге «Психология сознания» [Ревонсуо 2013] корректно и точно определил, как следует отвечать на вопрос о том, что такое сознание в методологическом смысле: «Описать сознание означает дать его определение, ввести понятия, которые ясно и системно описывают основные особенности сознания, привести показательные примеры этого феномена и дифференцировать его от других феноменов, с которыми его можно легко перепутать. Объяснить сознание означает связать его с другими феноменами, описав механизмы и принципы, лежащие в его основе или несущие ответственность за его возникновение, и показав, как сознание взаимодействует с мозгом и как оно руководит нашим поведением» [Ревонсуо 2013, 256]. Предлагаем найти аналог определения для понятия «нейрофилософия».

Эволюция термина «нейрофилософия»

Исторический анализ – один из основных методов исследования в философии. Полагаем, нет причин отказываться от него и в рамках нашего исследования. Главной отправной точкой нейрофилософии как отдельной области исследований стала книга П.С. Черчленд с одноименным названием «Нейрофилософия» [Churchland 1986]. Рассматриваемое в книге понимание нейрофилософии в значительной степени отличалось от его сегодняшней трактовки.

Нейрофилософия, описанная Черчленд, подана как объединение философии сознания и различных когнитивных наук, прежде всего нейронаук. С философской позиции Черчленд защищала идеи философии элиминативного материализма. С одной стороны, эта концепция предполагает, что сознание либо некоторые классы ментальных состояний не существуют. Они являются иллюзией, возникающей из-за наличия языка или долговременной памяти. С другой стороны, согласно данной концепции, не может существовать нейробиологического базиса для многих наших ментальных состояний (например, веры или желания). Подобных идей придерживались, например, бихевиористы.

Помимо идей элиминативизма Черчленд высказывала предположение о коэволюции психологии, нейробиологии и философии сознания до тех пор «пока в будущем, на некотором более высоком уровне, психологические теории не окажутся редуцированными к более фундаментальной нейрофизиологической теории; именно тогда возникнут предпосылки для разработки единой теории

сознания и мозга» [Философия… 2006, 365]. Под влиянием идей Черчленд сформулировано следующее определение нейрофилософии: это – «направление в современной философии науки, пытающееся обосновать правомерность редукции психологии к нейронаукам (нейробиологии, нейрофизиологии и нейропсихологии)» [Философия… 2006, 365]. Согласно трактовке, предложенной в энциклопедическом словаре под редакцией А.А. Ивина, одной из главных задач нейрофилософии является исследование возможностей компьютерного мышления, а также компьютерного моделирования природы мозга и сознания.

За прошедшие 20 лет количество оформившихся нейронаук необычайно расширилось. Например, появились социальная нейронаука, нейроархитектура, нейроэтика, нейроэкономика и т.д. Философско-методологическим базисом каждой из них должны быть конкретно-научные и философские теории. Общий базис должен быть назван философией нейронаук. Одновременно возникает вопрос о том, что означает нейрофилософия (нейронаучная философия).

Различие между нейрофилософией и смежными дисциплинами

В англоязычной литературе термины «философия нейронауки» и «нейрофилософия» сложно назвать синонимами. И. Голд и А.Л. Роскис пишут о том, что ряд исследователей различают философию нейронаук и нейрофилософию, перечисляя возможные направления философии нейронаук, не затрагивающие нейрофилософию: «анализ теоретических концепций; исследование методологий науки; отношение неврологии к другим наукам» [Gold, Roskies 2008, 350]. Они обращают внимание на то, что философия нейронауки мало распространена и что фактически лишь некоторые исследователи освещают эти вопросы. Поэтому указанные авторы предлагают в своей работе понимать «философию нейронаук» более широко, т.е. «как любое философское исследование, в котором нейронауки играют важную роль». Несмотря на четкое разделение терминов «философия нейронауки» и «нейрофилософия», иногда им сознательно пренебрегают.

М. Юнгерт придерживается более строгого различения. Он пишет: «В то время как философия нейронаук пытается применить методы и классические подходы философии науки к нейронаукам, т.н. нейрофилософия использует другой подход, применяя

результаты, полученные с помощью нейронаук, к классическим философским вопросам» [Jungert 2017]. В частности, он настаивает на том, что обсуждение объяснительных стратегий нейронаук относится к философии нейронаук, а разработка эмпирически обоснованных теорий относительно концепции морали или природы сознания – задача нейрофилософии. Подобного мнения придерживаются Дж. Бикл, П. Мандик и А. Ландрет. Они убеждены в том, что философия нейронауки занимается основополагающими проблемами нейронаук, а нейрофилософия пытается применить нейронаучные концепции к традиционным философским вопросам [Bickle, Mandik, Landreth 2019].

В итоге становится понятным, что если философию нейронаук рассматривают в первую очередь как философию науки, которая решает проблемы нейронаук, то под нейрофилософией понимают попытки применений различных концепций (к примеру, нейробиологических) к традиционным философским вопросам. Однако в русскоязычной литературе такое разделение встречается редко, поскольку возможность применения какой-либо науки в философии, как правило, считается традиционной проблемой философии этой науки. И наоборот: решение проблем нейронаук нередко основано на попытках совместить концепции нейронаук и философскую тематику.

Аналогичная ситуация происходит и при попытке разделить нейрофилософию и философию искусственного интеллекта, т.к. последняя часто не ограничивается методологической базой для попыток построения искусственного интеллекта и ставит более общие вопросы о том, «что такое интеллект», «что делает человека человеком», «где граница человеческого познания – там ли она, где границы познания машинного». Возможно поэтому в России обсуждение проблем и методов нейрофилософии в большей степени тесно связано, во-первых, с изучением работы мозга, во-вторых, с искусственным интеллектом.

Попытаемся выделить понятие нейрофилософии из ряда близких философских направлений. Согласно позиции ряда отечественных исследователей, под нейрофилософией следует понимать «конвенциональное обозначение стратегического направления философской науки, концентрированно характеризующего современные натуралистические интерпретации одного из “основных вопросов философии”» [Алексеев, Кузнецов, Савельев, Янковская 2015, 50]. Подчеркивается, что, несмотря на появив-

шуюся новую терминологию (среди новых терминов указаны субъективная реальность / нейрональная активность, когниция/ реализация), суть проблемы остается прежней. Речь снова идет о взаимоотношении психики и мозга, идеи и материи.

Новизна нейрофилософии заключается в первую очередь в натуралистическом тренде, который обусловлен развитием когнитивных наук и исследованиями в области искусственного интеллекта. А.Ю. Алексеев и его коллеги признают значение нейрофилософии как философии нейронаук, т.е. как философско-методологической рефлексии над основами нейронаук. Однако они дополняют представление о нейрофилософии определением, относящимся к иному направлению нейрофилософии: «Нейрофилософия – это систематическая форма изучения мировоззренческих аспектов, опирающаяся как на категориальные знания о нейрофизиологических основах психических явлений, так и на компьютерные методы имитации, моделирования, репродуцирования мозговой, психической и социальной активности» [Алексеев, Кузнецов, Савельев, Янковская 2015, 51].

Резюмируя, можно утверждать, что в отечественной философии выделено множество аспектов понимания «нейрофилософии»: философия науки, философия с акцентом на использование натурализма нейронаук (не только философия сознания, но и философия искусственного интеллекта, и социальная философия), элиминативный материализм, вариант философии сознания с опорой на нейронаучные теории и данные, а также междисциплинарное направление, изучающее системы мозга и системы, подобные мозгу.

В зависимости от того, какое понимание нейрофилософии нами использовано, на первый план выступает та или иная научная проблема. Перечислим ключевые проблемы, которые ученые пытаются решить в рамках нейрофилософии. Во-первых, проблема, связанная с определением того, что такое нейрофилософия и существуют ли у нее специфические характеристики. Во-вторых, психофизиологическая проблема сознания-тела. В-третьих, проблема построения искусственных нейросетей. В-четвертых, проблема методологии, которая включает в себя вопросы применения философских и метанаучных категорий в области нейронаук и, наоборот, применения результатов нейронаучных исследований в области философских и социальных исследований. И, наконец, актуален междисциплинарный

вопрос интеграции различных когнитивных и нейронаучных исследований.

Наличие одной проблемы для разных дисциплин приводит к тому, что провести разделение между такими дисциплинами непросто. Так, граница между нейрофилософией и философией сознания становится неясной. Причина очевидна: и нейронауки, и философия сознания рассматривают как ключевую проблему «сознание – мозг», но подход для ее решения у них различен. По мнению М. Юнгерта, разница между нейрофилософией и философией сознания существует только для философов, занимающихся исключительно философией сознания. Нейрофилософы не проводят строгой границы между эмпирическими, теоретическими и метатеоретическими вопросами, полагая, что нейрофилософия и философия сознания представляют собой единое направление исследований [Jungert 2017].

На наш взгляд, изложенный подход несколько радикализирован. Если нейронауки исследуют мозг, то нейрофилософия пытается искать общие подходы к его исследованию, а философия сознания исследует проблему «сознание – мозг» со стороны сознания. Чтобы продемонстрировать разделение между перечисленными дисциплинами более четко, рассмотрим две теории – теорию Эдельмана как пример нейронаучного подхода и концепцию коннекционизма в качестве синтеза идей философии нейронаук и философии искусственного интеллекта.

Теория Эдельмана как пример нейронаучного подхода

Ярким примером нейронаучного подхода может стать теория сознания Дж. Эдельмана, которую анализирует Д.И. Дубровский в статье «Нейрофилософия и проблема сознания» [Дубровский 2015], делая акцент на том, что Эдельман придает огромное значение философскому осмыслению своей теории и исследуемой проблемы. Охарактеризуем теорию Эдельмана подробнее и проследим связь его теории с философией.

Эдельман разрабатывает теорию отбора (селекции) нейрональных групп. Теория отбора нейрональных групп предложена им в рамках концепции нейродарвинизма. Ее сторонники настаивают на принципиальной изменчивости и приспосабливаемости нейронных сетей и групп. Эдельман утверждает, что при поступлении некоторого сигнала через органы последовательно активизируются разные этапы обработки этих сигналов. Каждый

из них производит своя группа нейронов. В дальнейшем группы, производящие обработку сигналов, объединяются и образуют систему нейрональных групп, что позволяет оптимизировать восприятие, т.е. ускорить или уточнить восприятие часто встречающихся сигналов.

Эдельман определяет первичное сознание как «состояние наличия ментальной осведомленности о вещах в мире» [Эдельман 2012, 422]. Он полагает, что первичное сознание является т.н. помнимым настоящим – *remembered present,* которое, в свою очередь, служит аналогом кажущегося настоящего – *specious present*, включающего в себя настоящее и краткий интервал недавнего прошлого. Высокоуровневое сознание, в отличие от первичного сознания, воплощает не только настоящее и недавнее прошлое, но и прошлое, и будущее. Оно относится прежде всего к прямому осознанию и сопровождается квалиами как формами высокоуровневой организации. Для построения модели первичного сознания, как правило, он применяет данные нейробиологии и физиологии. Однако представлено три пункта, где его теория тесно связана с философией.

Во-первых, теория Эдельмана построена на трех допущениях (физическом, эволюционном и квалиа). Физическое допущение предполагает, что законы физики не нарушаются, а духов и призраков, выходящих за рамки физического мира, не существует. Тем не менее физическое описание мира, согласно Эдельману, не является достаточным и полным. Эволюционное допущение сводится к убеждению о том, что сознание возникло в рамках эволюционного процесса, что оно реально действует и не является эпифеноменом. Квалиа допущение предусматривает, что квалии присущи всем человеческим существам и что не существует научных наблюдателей, которые были бы лишены квалий. Итак, все допущения предложены и сформулированы на базе ряда современных идей философии сознания.

Во-вторых, Эдельман разрабатывает исследовательскую программу и считает ее наиболее успешной. Он предлагает построение модели первичного сознания, надстроение над ней модели высокоуровневого сознания, а затем анализ связи построенных моделей с феноменальным опытом. Экспериментальное исследование, по его убеждению, должно происходить в обратном порядке: от феноменальных данных к нейрофизиологическому объяснению. Глубокий методологический подход ставит нас в

контекст методологии и философии науки, который Эдельман применяет исключительно к собственной теории.

Кроме того, он пытается решить такие традиционные философские вопросы, как существование квалий и проблему интерсубъективности. Все это позволяет утверждать, что его нейронаучная работа, хотя и не является примером теории философии сознания, но сопровождается философской и методологической работой.

Однако стоит обратить внимание и на следствие этой теории. Известно, что теория нейродарвизима Эдельмана стала основой теории селекции, применяющейся сегодня для ряда разработок из области искусственного интеллекта. В целом его теорию можно причислить к идеям, вдохновившим сторонников направления биовычислений (*Bio-inspired computing*) как метода создания искусственного интеллекта, который использует эволюционный подход, в отличие от традиционного, для искусственного интеллекта креационизма [Edelman 2007].

Целесообразно, как нам кажется, применить к этой концепции описанное выше разделение философии нейронауки и нейрофилософии. Философский анализ концепции Эдельмана, скорее, относится к философии науки, поскольку и его концепция как таковая является научной. Тем не менее поиск и анализ допущений и следствий из теории фактически относится к философии сознания, а проектирование исследовательской программы, как теоретической, так и экспериментальной, можно отнести к философии и методологии науки. Попытка же проанализировать вопрос о том, решает ли теория какие-то из традиционных проблем философии сознания, относится к сугубо философии сознания, несмотря на то, что используются богатые эмпирические данные физиологии и нейронаук. И, наконец, применение теории Эдельмана к проблемам искусственного интеллекта относится к философии искусственного интеллекта.

К нейрофилософии, таким образом, часто обращаются как к философии нейронаук, но подобный подход не совсем верен. Конечно, в рассматриваемой теории проблемы, связанные с искусственным интеллектом, можно легко отделить от остальных проблем. Однако в отношении нейрофилософии и философии нейронаук подобный подход неприемлем. Следует с осторожностью применять к понятию «нейрофилософия» термин «философия нейронауки». Разумнее отдавать предпочтение более точным и устоявшимся терминам «философия нейробиологии», «филосо-

фия когнитивных наук», «философия искусственного интеллекта» и др. Нейрофилософия в большей мере относится к философии сознания и отличается от иных разделов философии сознания тем, что она опирается более всего на нейронаучные теории и данные нейронаучных экспериментов.

Вместе с тем на примере теории Эдельмана становится очевидным тот факт, что разделить философию и методологию науки, философию искусственного интеллекта и философию сознания в полной мере не всегда возможно. Полноценный философский анализ нейронаучных теорий включает в себя и исследование методологии, и анализ исследовательской программы, и анализ философских допущений, находящийся в основе исследовательской программы, а также изучение возможных следствий и выводов из предлагаемых теорий. Фактически такое исследование допустимо считать междисциплинарным, затрагивающим и философию сознания, и когнитивные исследования, и искусственный интеллект. Если же сосредоточиться на методологических и метанаучных основах теории, то можно перейти в контекст философии науки и не использовать термин «нейрофилософия».

Коннекционизм как синтез идей философии нейронаук и философии искусственного интеллекта

Коннекционизм – это один из подходов к моделированию строения мозга, точнее, к моделированию биологических нейронных сетей. Суть главного принципа состоит в следующем: мышление как процесс может быть описано с помощью простых вычислительных элементов, которые связаны в сети. Как и большинство аксиом и принципов, принимаемых в любой науке, исследуемый принцип по своей сути относится к философии науки. Для биологической сети элементами являются нейроны, а связи – это синапсы. Для искусственной сети элементами служат, например, физические элементы памяти, а связями – провода, или все это может быть смоделировано в компьютерной программе.

Наиболее успешная реализация характеризуемого подхода – искусственные нейронные сети. Как правило, они состоят из большого количества соединенных элементов, разделенных на слои. Последние соответствуют трем классам: входной, нейроны которого получают информацию для обработки, выходной, нейроны которого выдают результаты обработки, и т.н. скрытый слой. Как замечает К. Бакнер, «если бы нейронная сеть модели-

ровала всю человеческую нервную систему, входные единицы были бы аналогичны сенсорным нейронам, выходные единицы – моторным нейронам, а скрытые – всем остальным нейронам» [Buckner 2019].

Между нейронами существуют связи, которые имеют вес активации. Сигнал от слоя нижележащих нейронов подается по линиям связи в слой вышележащих. Этот сигнал умножается на число веса активации, и, если суммарное значение от всех входящих линий превышает порог активации нейрона, он переходит в возбужденное состояние и посылает сигнал дальше. В ином случае распространение сигнала блокируется. Обучить сеть – значит подобрать значения веса активации таким образом, чтобы минимизировать расхождение между входными и выходными данными.

Наиболее интересна интерпретация термина «информация», которая в описанной системе представлена весами связей между нейронами. Она сосредоточена не в нейроне, а в связях между ними, т.е. имеет не локальный, а дистрибутивный характер. Среди задач, решаемых искусственной нейросетью, – распознавание образов, кластеризация (разбиение множества входных сигналов на классы), прогнозирование, аппроксимация, сжатие данных и т.д. Методологическая база изложенного подхода относится к философии сознания и представлена теорией функционализма, теориями сознания высших порядков. Функционализм интерпретирует сознание как сложную функцию между входными данными для сенсорных нейронов и поведением, интерпретируемым как выходные данные моторных нейронов. Такая теория, по сути, убирает понятия сознания и самосознания как необходимые для процесса моделирования. Мозг интерпретирован как процессор, а психология становится теорией о том, какие операции нужно произвести, чтобы добиться нужного поведения.

Коннекционистская модель в настоящее время применяется и в обратную сторону – к изучению мозга. С. Сеунг отмечает: «Коннекционисты рассматривают зоны мозга не как некую элементарную ячейку, а как сложную сеть, которая состоит из множества нейронов» [Сеунг 2014, 20]. Теоретически в мозге происходят следующие процессы, если рассматривать его с точки зрения коннекционистской модели: изменение удельного веса существующих связей, рекомбинация, переподключение, реге-

нерация [Сеунг 2014, 14]. Только первое свойство из этого списка характерно и для искусственных нейросетей. Остальные свойства обеспечены биологической природой мозга и, по-видимому, могут быть сведены к такому свойству, как пластичность.

Сегодня искусственные нейросети обладают рядом проблем, наличие которых препятствует видеть в них слабый искусственный интеллект. Самая главная из проблем – отсутствие понимания, т.е. у системы нет семантической интерпретации данных, которые подаются ей на вход. В более общем смысле проблема связана с тем, что в нейросети не запрограммировано понятие «здравого смысла» и концепции общего устройства мира. Иными словами, если биологическое сознание работает с символами и следует причинно-следственным связям, то искусственная нейросеть работает с числами и следует закономерностям. Такая проблема не принадлежит частным наукам – она сугубо философская, поэтому может быть отнесена к философии искусственного интеллекта и, на наш взгляд, решение следует ожидать в этой области.

Другая проблема – тонкие надстройки, которые могут привести к неверной классификации изображений. Например, добавление «шума» к изображению лица способно помешать работе программ распознавания лиц. Речь идет уже о конкретно-научной проблеме. Она описана, к примеру, в одной из недавних статей под названием «Слон в комнате» [Rosenfeld, Zemel, Tsotsos 2018]. Предложена гипотетическая ситуация: классификация объектов, находящихся в комнате. Дан набор фотографий. На некоторые из них посредством программ обработки изображений поместили уменьшенное изображение реального слона. При этом программа распознавания образов могла не замечать слона во многих местах или классифицировать его неверно, или неправильно классифицировать объекты, расположенные рядом.

На примере коннекционизма можно увидеть, где проходит разделение между философией науки, философией нейронаук и философией искусственного интеллекта. Базой подхода являются концепции философии науки в целом и философии сознания в частности. И рефлексия над подходом, и рефлексия проблем – это философия искусственного интеллекта. Ввиду приведенного примера становится очевидным, что любое фундаментальное теоретическое исследование по своей сути всегда будет междисциплинарным.

Заключение

Подведем итоги нашего исследования. Представление о нейрофилософии исключительно как о попытке оправдания элиминативного материализма устарело и не соответствует действительности. Представление о нейрофилософии как о философии нейронаук не вполне корректно: поскольку под нейрофилософией понимается целый ряд направлений, термин «нейрофилософия» становится чрезмерно нагруженным, и каждый раз при его использовании необходимо уточнять, что имеется в виду. Предлагаем разграничить три термина.

Нейрофилософия – это направление в философии начала XXI века, использующее нейронаучные концепции для решения традиционных философских проблем. Философия нейронаук, по нашему мнению, – чрезвычайно абстрактный термин, который содержит исследование научно-философских концепций, лежащих в основе той или иной нейронауки, а также изучение и анализ методов, проблем и целей отдельных нейронаук. Таким образом, философия нейронаук может быть рассмотрена в первую очередь как раздел философии науки, который формулирует и решает проблемы нейронаук. Более корректен в методологическом аспекте отказ от этого понятия в пользу таких терминов, как философия нейроэтики, философия нейробиологии. Философия искусственного интеллекта – это философско-методологическая база для изучения небиологического интеллекта. В таком понимании предложенное определение находится в отношениях пересечения с философией нейронаук.

Методологии науки известно, что любое исследование проходит два этапа на своем пути. Первый этап – анализ и дифференциация гипотез, касающихся объяснений изучаемого факта. Второй этап – синтез общих и отличающихся моментов в гипотезах, который позволяет создать абстрактную обобщенную теорию изучаемого явления. В становлении нейронаук и их научно-философского базиса наблюдается все еще пребывание на первом этапе. Философия нейронаук как база всех существующих нейронаучных теорий возникнет через определенное время, и именно в соответствующий период будет необходим данный термин.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Алексеев, Кузнецов, Савельев, Янковская 2015 – *Алексеев А.Ю., Кузнецов В.Г., Савельев А.В., Янковская Е.А.* Становление отечественной нейрофилософии // Философские науки. 2015. № 11. С. 48–66.

Дубровский 2015 – *Дубровский Д.И.* Нейрофилософия и проблема сознания // Философские науки. 2015. № 11. С. 9–22.

Ревонсуо 2013 – *Ревонсуо А.* Психология сознания. – СПб.: Питер, 2013.

Сеунг 2014 – *Сеунг С.* Коннектом. Как мозг делает нас тем, что мы есть. – М.: БИНОМ. Лаборатория знаний, 2014.

Философия… 2006 – Философия. Энциклопедический словарь / под ред. А.А. Ивина. – М.: Гардарики, 2006.

Эдельман 2012 – *Эдельман Дж.* Сознание: помнимое настоящее // Эволюционная эпистемология. Антология. – М.; СПб.: Центр гуманитарных инициатив, 2012. С. 419–438.

Bickle, Mandik, Landreth 2019 – *Bickle J., Mandik P., Landreth A.* The Philosophy of Neuroscience // Stanford Encyclopedia of Philosophy / ed. E.N. Zalta. 2019. – URL: https://plato.stanford.edu/archives/fall2019/entries/neuroscience

Buckner 2019 – *Buckner C.* Connectionism // Stanford Encyclopedia of Philosophy / ed. by E.N. Zalta. 2019. – URL: https://plato.stanford.edu/entries/connectionism

Churchland 1986 – *Churchland P.S.* Neurophilosophy: Toward a Unified Science of the Mind-Brain (Computational Models of Cognition and Perception). – Cambridge, MA: MIT Press, 1986.

Edelman 2007 – *Edelman G.M.* Learning in and from Brain-Based Devices // Science. Vol. 318. No. 5853. P. 1103–1105.

Gold, Roskies 2008 – *Gold I., Roskies A.L.* Philosophy of Neuroscience // The Oxford Handbook of Philosophy of Biology / ed by M. Ruse. – Oxford: Oxford University Press, 2008. P. 349–380.

Jungert 2017 – *Jungert M.* Neurophilosophy or Philosophy of Neuroscience? What Neuroscience and Philosophy Can and Cannot Do for Each Other // The Human Sciences after the Decade of the Brain. Perspectives on the Neuro-Turn in the Social Sciences and the Humanities / ed. by E. Hildt, J. Leefmann. – London: Academic Press, 2017. P. 3–13.

Rosenfeld, Zemel, Tsotsos 2018 – *Rosenfeld A., Zemel R., Tsotsos J.* The Elephant in the Room // ArXiv. – URL: https://arxiv.org/abs/1808.03305

REFERENCES

Alekseev A.Y., Kuznetsov V.G., Saveliev A.V., & Yankovskaya E.A. (2015) The Formation of the National Neurophilosophy. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* No. 11, pp. 48–66. (In Russian).

Bickle J., Mandik P., & Landreth A. (2019) The Philosophy of Neuroscience. In: Zalta E.N. (Ed.) *The Stanford Encyclopedia of Philosophy.* Retrieved from https://plato.stanford.edu/archives/fall2019/entries/neuroscience

Buckner C. (2019) Connectionism. In: Zalta E.N. (Ed.) *The Stanford Encyclopedia of Philosophy.* Retrieved from https://plato.stanford.edu/entries/connectionism

Churchland P.S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind-Brain (Computational Models of Cognition and Perception).* Cambridge, MA: MIT Press.

Dubrovsky D.I, (2015) Neurophilosophy and the Problem of Consciousness. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* No. 11, pp. 9–22 (in Russian).

Edelman G.M (2001) Consciousness: The Remembered Present. *Annals of the New York Academy of Sciences. Vol. 929, no.* 1, pp. 111–122 (Russian translation in: Kniazeva E.N. (Ed.) *Evolutionary Epistemology. Anthology* (pp. 419–438). Moscow; Saint Petersburg: Tsentr gumanitarnykh initsiativ, 2012).

Edelman G.M. (2007) Learning in and from Brain-Based Devices. *Science.* Vol. 318, no. 5853, pp. 1103–1105.

Gold I. & Roskies A.L. (2008) Philosophy of Neuroscience. In: Ruse M. (Ed.) *The Oxford Handbook of Philosophy of Biology* (pp. 349–380). Oxford: Oxford University Press.

Ivin A.A. (Ed.) (2004) *Philosophy: Encyclopedic Dictionary.* Moscow: Gardariki. (In Russian).

Jungert M. (2017) Neurophilosophy or Philosophy of Neuroscience? What Neuroscience and Philosophy Can and Cannot Do for Each Other. In: Hildt E. & Leefmann J. (Eds.): *The Human Sciences after the Decade of the Brain. Perspectives on the Neuro-Turn in the Social Sciences and the Humanities* (pp. 3–13). London: Academic Press.

Revonsuo A. (2009) *Consciousness: The Science of Subjectivity.* Hove: Psychology Press (Russian translation: Saint Petersburg: Piter, 2013).

Rosenfeld A., Zemel R., & Tsotsos J. (2018) *The Elephant in the Room.* arXiv. Retrieved from https://arxiv.org/abs/1808.03305

Seung S. (2012) *Connectome: How the Brain's Wiring Makes Us Who We Are.* Boston: Houghton Mifflin Harcourt (Russian translation: Moscow: BINOM, 2014).

Векторы развития искусственного интеллекта

DOI: 10.30727/0235-1188-2021-64-1-88-101
Оригинальная исследовательская статья
Original research article

Искусственный интеллект в субъектных парадигмах управления*

В.Е. Лепский
Институт философии РАН, Москва, Россия

Аннотация

При исследовании возможностей и ограничений использования искусственного интеллекта (ИИ) в процессах управления возникает много нерешенных проблем, которые в значительной степени связаны с тем, что представления об ИИ формируются без учета парадигм управления. Наиболее распространенными являются парадигмы, рассматривающие искусственный интеллект не как средства, включенные в управленческую деятельность и соответствующие парадигмы управления, а как самостоятельные объекты исследования в соответствующих специфике этих объектов парадигмах. Эти парадигмы ИИ способствуют развитию отдельных направлений проблематики ИИ, но затрудняют использование таких разработок в процессах управления и оставляют без внимания многие потенциальные направления представлений об ИИ, актуальные для развития проблематики управления. Организация процессов управления базируется на своих специализированных парадигмах (субъектных, кибернетических и др.), выдвигающих специфические требования к представлениям об ИИ, а также задачам, в которых целесообразно использовать ИИ. В этих парадигмах управления формируется прямой заказ к ИИ, что способствует успешному практическому использованию и развитию ИИ, а также созданию механизмов контроля и нейтрализации негативных последствий. Автором предложен механизм взаимодействия субъектов (людей) и активных формы ИИ (рассматриваемых в качестве псевдосубъектов). С учетом тренда повышения роли рефлексивной активности в процессах социального управления в статье рассмотрены место и роль ИИ в обеспечении рефлексивной активности в субъект-

* Работа выполнена при поддержке РФФИ проект № 20-511-00003 «Философский, методологический и междисциплинарный анализ перспектив и рисков развития цифровой реальности».

ных парадигмах управления. Анализ трендов развития проблематики управления с позиций развития научной рациональности (классической, неклассической, постнеклассической) позволяет сделать вывод о том, что каждая субъектная парадигма управления («субъект – объект», «субъект – субъект», «субъект – метасубъект (саморазвивающаяся полисубъектная среда)») имеют свою специфику, которая должна учитываться при разработке активных форм ИИ.

Ключевые слова: философия искусственного интеллекта, научная рациональность, субъект, процессы управления, рефлексия, рефлексивная активность, кибернетика.

Лепский Владимир Евгеньевич – доктор психологических наук, главный научный сотрудник сектора междисциплинарных проблем научно-технического развития Института философии РАН.

VELepskiy@mail.ru

https://orcid.org/0000-0002-6893-0234

Для цитирования: *Лепский В.Е.* (2021) Искусственный интеллект в субъектных парадигмах управления // Философские науки. 2021. Т. 64. № 1. С. 88–101. DOI: 10.30727/0235-1188-2021-64-1-88-101

Artificial Intelligence in Subject-Oriented Control Paradigms*

V.E. Lepskiy

Institute of Philosophy, Russian Academy of Science, Moscow, Russia

Abstract

The article discusses unresolved problems and limitations that arise with application of artificial intelligence (AI). These problems are largely related to the fact that ideas about AI are often formed without taking into account the control paradigms. The most common ones are paradigms that consider artificial intelligence not as means included in control activities or control paradigms, but as independent objects of research in the paradigms corresponding to the specifics of such objects. Such paradigms contribute to the development of certain areas of AI, but they also complicate their application in control processes and ignore many potential areas of AI that are relevant to the development of control problems. The organization of

* The research was supported by the grant of the Russian Foundation for Basic Research (RFBR), project no. 20-511-00003 "Philosophical, Methodological, and Multidisciplinary Analysis of Prospects and Risks of the Development of Digital Reality."

control processes is based on their specific paradigms (subjective, cybernetic, etc.) that set such specific requirements to AI implementations as well as to tasks in which it is advisable to use AI. Such control paradigms form tasks for AI, which contributes to successful practical application and development of AI as well as to mechanisms for controlling and neutralizing negative consequences. The author proposes a mechanism for interaction of subjects (persons) and active forms of AI (considered as pseudo-subjects). Taking into account the increasing role of reflexive activity in the processes of social control, the article considers the place and role of AI in ensuring reflexive activity in the subject paradigms of control. Analysis of trends in the development of controlling from the standpoint of the development of scientific rationality (classical, non-classical, and post-non-classical) allows us to conclude that each subject paradigm of control ("subject – object," "subject – subject," and "subject – meta-subject") has its own specifics, which should be considered when developing active forms of AI.

Keywords: philosophy of artificial intelligence, scientific rationality, subject, control processes, reflection, reflection activity, cybernetics.

Vladimir E. Lepskiy – D.Sc. in Psychology, Chief Research Fellow, Department of Interdisciplinary Problems in the Advance of Science and Technology, Institute of Philosophy, Russian Academy of Science.
VELepskiy@mail.ru
http://orcid.org/0000-0002-6893-0234

For citation: Lepskiy V.E. (2021) Artificial Intelligence in Subject-Oriented Control Paradigms. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 64, no. 1, pp. 88–101.
DOI: 10.30727/0235-1188-2021-64-1-88-101

Введение

При исследовании возможностей и ограничений применения искусственного интеллекта (ИИ) в процессах управления возникает немало проблемных вопросов, связанных с разрывом парадигм, задающих представления об ИИ вне учета парадигм управления [Лепский 2020]. В большей степени распространены парадигмы, рассматривающие искусственный интеллект не как средства, включенные в управленческую деятельность и соответствующие парадигмы управления, а как самостоятельные объекты исследования в соответствующих специфике этих объектов парадигмах.

Приведем примеры таких парадигм искусственного интеллекта:

– *морфологическая парадигма* – представление об ИИ как о модели мозга;

– *логическая парадигма* – представление об ИИ как о решателе задач;

– *нейрокибернетическая парадигм*а – представление об ИИ как о нейросетях;

– *имитационная парадигма* – подобие человеческим рассуждениям (подход Тьюринга);

– *парадигма слабого ИИ* – моделирование видов человеческой деятельности, которые традиционно считаются интеллектуальными;

– *парадигма сильного ИИ* предполагает, что компьютеры могут приобрести способность мыслить и осознавать себя как отдельную личность, хотя и не обязательно, что их мыслительный процесс будет подобен человеческому;

– *парадигма общего ИИ*, способного самообучаться и решать разнообразные задачи в разных контекстах [Сильный искусственный интеллект… 2021].

Эти парадигмы ИИ способствуют развитию отдельных направлений проблематики ИИ, но затрудняют использование этих разработок в процессах управления, оставляя без внимания многие потенциальные направления представлений об ИИ, актуальные для развития проблематики управления. Организация процессов управления базируется на своих специализированных парадигмах (субъектных, кибернетических и др.), выдвигающих специфические требования к представлениям об ИИ, а также задачам, в которых целесообразно применять ИИ.

Фактически речь идет о том, что для использования ИИ в процессах управления необходима организация двух уровней технологического обеспечения, включающего в себя ИИ. Первый уровень, концептуально-технологический, непосредственно связан с парадигмами управления, в которых проведено соотнесение со специализированными технологиями ИИ. Второй уровень – инструментально-технологический. В него включены разработки ИИ на основе любых парадигм. Отдельная проблема – установление интерфейса между этими двумя уровнями в интересах развития проблематики управления и проблематики ИИ.

Полагаем, что в качестве актуальной проблемы развития искусственного интеллекта выступает философско-методологическая проблема формирования представлений об искусственном интел-

лекте в контексте парадигм конкретных видов человеческой деятельности. В данном случае – в контексте парадигм управления. Анализ трендов развития проблематики управления с позиций развития научной рациональности (классической, неклассической, постнеклассической) позволяет сделать вывод о том, что они в значительной степени связаны с развитием соответствующих субъектных парадигм управления [Лепский 2019; Lepskiy 2018].

Предлагаем представление об ИИ в субъектных парадигмах управления, соответствующих эволюции представлений о научной рациональности.

Субъектные парадигмы в управлении

На базе современных представлений философии науки [Степин 2003, 619–636] в настоящее время предложено системное видение эволюции проблематики управления, которое может стать основой понимания места и роли ИИ в совершенствовании механизмов управления различными типами систем, включая и социальные системы. Обоснована связь базовых этапов развития кибернетики с этапами развития представлений о научной рациональности [Лепский 2019].

Классической кибернетике (Н. Винер) соответствует этап становления классической научной рациональности. Доминирующее философское обеспечение проблематики управления выполнял *позитивизм.* Доминирующая субъектная парадигма «*субъект – объект*» [Лекторский 2001] отражает суть классической кибернетики, т.е. кибернетики «наблюдаемых систем». Доминирующим подходом в обеспечении деятельностной активности выступал *деятельностный подход* (А.Н. Леонтьев), в котором роль субъекта существенно ограничена целями и нормами деятельности.

Кибернетика второго порядка (Ф. Ферстер) соответствует этапу становления неклассической научной рациональности. Доминирующее философское обеспечение проблематики управления выполнял *философский конструктивизм*. Доминирующая субъектная парадигма «*субъект – субъект*» отражает суть кибернетики второго порядка, т.е. кибернетики «наблюдающих систем». Субъект управления трактуется как один из участников процессов управления, в которых активную роль играет и объект управления. Последний также рассматривается как субъект [Lefebvre 1986]. Доминирующим подходом в обеспечении деятельностной, коммуникативной и рефлексивной активности становит-

ся *субъектно-деятельностный подход* [Рубинштейн 1997, 438], в котором роль субъекта существенно возрастает.

Становление кибернетики третьего порядка [Lepskiy 2018] целесообразно связывать со становлением этапа постнеклассической научной рациональности [Степин 2003, 619–636]. Доминирующее философское обеспечение проблематики управления выполняла *гуманистическая трактовка философского конструктивизма* [Лепский 2019]. В центре внимания проблематики управления оказываются саморазвивающиеся системы [Степин 2003, 628–629]. Доминирующей субъектной парадигмой в управлении становится парадигма «*субъект – метасубъект (саморазвивающаяся среда)*». В качестве метасубъекта рассматривается саморазвивающаяся полисубъектная система (среда), наделенная свойствами субъектности, в которую погружен и с которой может взаимодействовать через рефлексию субъект. В качестве метасубъекта может выступать семья, организация, страна, человечество [Лепский 2010].

Доминирующим подходом в обеспечении деятельностной, коммуникативной и рефлексивной активности становится *субъектно-ориентированный подход* [Лепский 2019], в котором в центре внимания оказывается субъект и его взаимодействия с другими субъектами и метасубъектами. Обоснование формирования субъектно-ориентированного подхода, интегрирующего деятельностный и субъектно-деятельностный подходы, обусловлено необходимостью рассмотрения широкого спектра связей субъектов с культурой, а также снижением роли нормативных регуляторов деятельности и коммуникаций.

Активные формы искусственного интеллекта как псевдосубъекты

В контексте субъектных парадигм управления для обеспечения разнообразных взаимодействий субъектов естественного интеллекта с образованиями ИИ их целесообразно трактовать в качестве активных форм ИИ, обладающих базовыми свойствами субъектов, инвариантными по отношению к типу субъекта (индивида, группы, организации и др.), в состав которого входят носители естественного интеллекта. Фактически речь должна идти об псевдосубъектах ИИ.

К базовым инвариантным свойствам субъектов целесообразно отнести целеустремленность, рефлексивность, коммуникативность, социальность и способность к развитию [Лепский 2019].

Эти характеристики могут быть интерпретированы как характеристики псевдосубъектов ИИ, что находит отражение в многочисленных публикациях по кибернетике и ИИ.

Характеристика «рефлексивность» занимает особое место среди всех характеристик псевдосубъектов. Во-первых, рефлексивность положена в основу выделения этапов развития проблематики управления (кибернетика первого, второго и третьего порядка). Во-вторых, она обеспечивает все другие характеристики субъектов и псевдосубъектов. В-третьих, она способствует целостности рефлексивных субъектов и псевдосубъектов, механизмов их сборки. В-четвертых, рефлексивность находится в основе механизмов обеспечения идентичности по отношению к макросубъектам.

Модели псевдосубъектов ИИ должны эволюционировать в совершенствовании псевдохарактеристик субъектности, приближаясь по мере развития науки к идеалам субъектов на базе естественного интеллекта.

Совместное сосуществование субъектов естественного интеллекта и псевдосубъектов ИИ образует гибридную среду. В этой среде происходят как взаимодействия разнородных образований, так и их интеграция в новые виды образований на основе субъектов и псевдосубъектов. В этой связи целесообразно ввести для субъектов и псевдосубъектов общее понятие активных элементов (АКТЭЛов).

Под АКТЭЛами будем понимать активные элементы, обладающие базовыми инвариантными характеристиками субъектов, псевдосубъектов, реализованные на субстратном уровне естественным или искусственным интеллектом, а также их комбинациями. Это понятие предложено и использовано автором для проектирования автоматизированных систем организационного управления [Лепский 1998].

Понятие «АКТЭЛ» шире часто используемого в управленческой практике понятия «Агент». Агенту передаются от субъектов или псевдосубъектов определенные полномочия выполнения функций. АКТЭЛ может быть Агентом, но может быть и самостоятельным субъектом или псевдосубъектом. Целесообразность применения понятия «АКТЭЛ» связана с организацией сред гибридной реальности, состоящих из носителей естественного и искусственного интеллекта, а также интегрированных образований естественного и искусственного интеллекта.

Механизмы взаимодействия субъектов и активных форм искусственного интеллекта

Одним из базовых свойств субъекта является рефлексия. Базовый принцип организации взаимодействий и интеграции субъектов и псевдосубъектов (АКТЭЛов) в средах гибридной реальности естественного и искусственного интеллекта должен определять процедуру их рефлексивного согласования. Для понимания специфики саморегуляции рефлексивной активности наибольшее значение имеет принцип «двойного субъекта» [Лепский 1998]. Суть данного принципа состоит в совершенствовании различных видов активности (деятельностной, коммуникативной, рефлексивной) субъектов (псевдосубъектов) за счет организации взаимодействия с партнерами, имеющими адекватные позиции, в формировании таких партнеров, передаче другим субъектам (псевдосубъектам) устоявшихся видов активности, организации на основе анализа активности субъектов их двойников (моделей целостных субъектов).

Принцип двойного субъекта для организации сред поддержки принятия управленческих решений сформулирован более 20 лет назад [Лепский 1998]. Этот принцип определяет технологические процедуры взаимодействия и интеграции АКТЭЛов в средах гибридной реальности [Лепский 2019].

Отдельные аспекты представлений о принципе двойного субъекта отражены в широком спектре междисциплинарных исследований: теории функциональных систем П.К. Анохина, теории коррективного управления движениями Н.А. Бернштейна, учении о доминанте и хронотопе А.А. Ухтомского, модели двигательной активности В.П. Зинченко, организации систем искусственного интеллекта Д.А. Поспелова, рефлексивном критерии цивилизации В.А. Лефевра [Лепский 1998].

Есть основания полагать, что учет принципа двойного субъекта в организации систем с использованием ИИ позволит сделать очередной шаг к созданию систем с включением элементов ИИ, не выходящих из-под контроля их создателей. Принцип двойного субъекта также дает основания для утверждения о важнейшей роли рефлексивной активности при рассмотрении ИИ в субъектных парадигмах управления [Lefebvre 1986; Лепский 2017].

Искусственный интеллект в обеспечении рефлексивной активности

Парадигма «субъект – объект»

В субъектной парадигме «*субъект – объект*» доминирует деятельностная активность как следствие рефлексивной активности, и этому в управленческой практике отводится второстепенная роль. Стимулирование и поддержка рефлексивной активности оказывается актуальной в нестандартных случаях для обеспечения процессов принятия решений в проблемных ситуациях. Их разрешение требует выхода субъекта за рамки сложившейся управленческой деятельности, требует «рефлексивного выхода». Проблема «рефлексивного выхода» привлекла внимание методологов [Щедровицкий 1975] и психологов. Психологи (В.А. Петровский, Я.А. Пономарев, Д.Б. Богоявленская, А.Г. Асмолов, И.Н. Семенов и др.) обозначили ее как проблему надситуационной активности.

Для управленческой практики данная проблема представляет интерес в двух аспектах. Во-первых, при выявлении и совершенствовании способности управленцев к надситуационной активности. Во-вторых, при разработке механизмов стимулирования и поддержки надситуационной активности. Второй аспект может вызывать интерес при постановке и решении новых сложных задач с использованием ИИ. Для их решения могут применяться активные формы ИИ, которые должны обеспечить создание псевдосубъектного партнера для субъекта управления, помогающего оценивать сложившиеся ситуации точек разрыва управленческой деятельности, предлагающего либо имеющиеся решения, либо стимулирующие рефлексивный выход для поиска нестандартных решений и поддержку в условиях рефлексивного выхода. Следует отметить, что для постановки и решения этих задач существует богатый междисциплинарный научный задел [Лепский 1998].

Рефлексивная активность объекта управления в рамках субъектной парадигмы «*субъект – объект*» остается вне рассмотрения.

Парадигма «субъект – субъект»

В субъектной парадигме «*субъект – субъект*» наряду с деятельностной активностью резко возрастает роль коммуникативной активности и обеспечивающих ее форм рефлексивной активности (коммуникативной рефлексии). Это проявилось в переходе от

деятельностного подхода к субъектно-деятельностному подходу [Рубинштейн 1997].

Интенсивные исследования возможностей использования рефлексивной активности в процессах управления, на примере конфликтных ситуаций, проведены в 60–70-е годы прошлого столетия в СССР под руководством В.А. Лефевра [Лефевр 1973]. Разработаны математические модели и выполнены экспериментальные исследования. Предложено понятие рефлексивного управления, которое нашло широкое применение в военной сфере, педагогических технологиях, экономике и др. [Novikov 2016].

В контексте проблематики ИИ нельзя не отметить, что созданы автоматы (псевдосубъекты), которые с учетом стереотипных форм рефлексивной активности человека могли переигрывать его в конфликтных взаимодействиях, даже не зная цели человека (например, стороны платежной матрицы в моделях теории игр с нулевой суммой) [Лефевр 1973].

Рефлексивное управление – одна из форм рефлексивных технологий, включающих в себя широкий спектр технологий управления рефлексивной активностью субъектов и псевдосубъектов. Предложено три группы рефлексивных технологий: имитация рефлексивной активности, рефлексивное управление, рефлексивное программирование [Лепский 2016, 73–78]. Эти рефлексивные технологии могут использоваться как в конфликтных ситуациях, так и для поддержки совместной деятельности в разнообразных сферах деятельности [Kauffman 2016; Espejo 2015; Müller 2015].

Использование ИИ возможно в обеспечении всех рефлексивных технологий за счет создания адекватных видов псевдосубъектов ИИ. Фактически речь может идти о создании сообществ псевдосубъектов, соответствующих рефлексивным структурам конкретных рефлексивных технологий.

Примером возможности использования системы псевдосубъектов ИИ для обеспечения рефлексивной активности может служить поддержка принятия решений в условиях переговорных процессов, когда партнер использует 36 китайских стратагем. Рефлексивный анализ 36 китайских стратагем позволил поставить в соответствие каждой стратагеме определенную рефлексивную технологию. Это дает возможность за счет усложнения рефлексивной технологии, адекватной используемой стратагеме, нейтрализовать угрозу манипулятивных воздействий и перевести взаимодействие из конфликта в партнерские отношения

[Лепский 2016, 82–88]. На практике возникает проблема оперативного прогнозирования и идентификации применяемой стратагемы. Данная проблема могла бы успешно решаться с использованием системы специализированных псевдосубъектов ИИ, которые оперативно осуществляли бы идентификацию стратагем и советовали применять адекватные рефлексивные технологии на основе принципа двойного субъекта.

Отметим, что сфера образования могла бы стать одной из самых актуальных для применения ИИ в поддержке различных форм рефлексивной активности в субъектной парадигме «*субъект – субъект*».

Парадигма «*субъект – метасубъект*»

Субъектная парадигма «*субъект – метасубъект* (*саморазвивающаяся среда*)» формируется в контексте постнеклассической научной рациональности, в которой в центре внимания оказываются саморазвивающиеся человекоразмерные системы [Степин 2003, 619–636]. Важно отметить, что управленческие науки находятся под влиянием как внутринаучных, так и вненаучных (социальных) ценностей. Саморазвивающиеся человекоразмерные системы при этом органично вписаны в культуру.

Эти соображения дают основания для введения понятия саморазвивающихся полисубъектных (рефлексивно-активных) сред, которые рассматриваются как целостные образования – метасубъекты. С ними могут идентифицировать себя входящие в них субъекты (псевдосубъекты). Субъектные аспекты в таких средах возросли, что стимулировало становление субъектно-ориентированного подхода, обеспечивающего конвергенцию с деятельностным и субъектно-деятельностным подходами.

Метасубъект является виртуальным образованием, что способствует повышению роли рефлексивной активности в субъектной парадигме «*субъект – метасубъект* (*саморазвивающаяся среда*)». Появляется новая форма рефлексивной активности – метасубъектная рефлексия.

Бытие субъектов и псевдосубъектов в саморазвивающихся полисубъектных (рефлексивно-активных) средах задается системой онтологий, которая обеспечивает сборку в целое (метасубъекта) включенных в среду субъектов (псевдосубъектов). Нами разработана система онтологий, в которую входят онтологии обеспечения жизнедеятельности, преодоления точек разрыва жизнедеятельности, стратегического целеполагания, разработки стратегий и про-

ектов, внедрения и инновационного обеспечения стратегий и проектов. Эта система онтологий прошла научно-методологическую и практическую апробацию [Лепский 1998; Lepskiy 2018].

Постановка задач для ИИ должна осуществляться через систему онтологий бытия субъектов и псевдосубъектов, а также в контексте поддержки рефлексивной активности субъектов и псевдосубъектов по отношению к метасубъектам.

В настоящее время делаются первые шаги в осмыслении роли и задач ИИ в субъектной парадигме «*субъект – метасубъект* (*саморазвивающаяся среда*)». Но есть основания утверждать, что разработанные философско-методологические основания позволят интенсифицировать разработки ИИ в интересах совершенствования процессов управления и развития социальных систем [Umpleby, Medvedeva, Lepskiy 2019].

Заключение

Организация процессов управления базируется на специализированных парадигмах (субъектных, кибернетических и др.), выдвигающих специфические требования к представлениям об ИИ и задачам, в которых целесообразно использовать ИИ. Каждому типу научной рациональности (классическая, неклассическая, постнеклассическая) соответствует своя субъектная парадигма в управлении («субъект-объект», субъект-субъект», субъект-метасубъект»).

Активная форма ИИ приобретает субъектность и потому может рассматриваться как псевдосубъект. Организацию механизмов взаимодействия людей (субъектов) и активных форм ИИ (псевдосубъектов) предлагается построить на основе принципа «двойного субъекта», позволяющего совершенствовать различных видов активности за счет организации взаимодействия с партнерами и делегации им определенных задач. Для каждой субъектной парадигмы управления («субъект – объект», «субъект – субъект», «субъект – метасубъект») характерен свой доминирующий вид активности (соответственно – деятельностная коммуникативная, рефлексивная активность). Для совершенствования возможностей ИИ ключевую роль играет рефлексивная активность. Поэтому нами были сформулированы базовые задачи рефлексивной активности ИИ в каждой из трех субъектных парадигм управления.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Лекторский 2001 – *Лекторский В.А.* Эпистемология классическая и неклассическая. – М.: Эдиториал УРСС, 2001.

Лепский 1998 – *Лепский В.Е.* Концепция субъектно-ориентированной компьютеризации управленческой деятельности. – М.: Институт психологии РАН, 1998.

Лепский 2010 – *Лепский В.Е.* Рефлексивно-активные среды инновационного развития. – М.: Когито-Центр, 2010.

Лепский 2016 – *Лепский В.Е.* Технологии управления в информационных войнах (от классики к постнеклассике) – М.: Когито-Центр, 2016.

Лепский 2017 – *Лепский В.Е.* Рефлексивная активность в управлении // Институт психологии Российской академии наук. Человек и мир. 2017. Т. 1. № 1. С. 53–80.

Лепский 2019 – *Лепский В.Е.* Методологический и философский анализ развития проблематики управления. – М.: Когито-Центр, 2019.

Лепский 2020 – *Лепский В.Е.* Философско-методологические основания оценки социально-психологических последствий внедрения новых технологий // Психологический журнал. 2020. Т. 41. № 4. С. 105–108.

Лефевр 1973 – *Лефевр В.А.* Конфликтующие структуры. – М.: Советское радио, 1973.

Рубинштейн 1997 – *Рубинштейн С.Л.* Избранные философско-психологические труды. Основы онтологии, логики и психологии. – М.: Наука, 1997.

Сильный искусственный интеллект… 2021 – Сильный искусственный интеллект. На подступах к сверхразуму / А.А. Ведяхин и др.; науч. ред. А.С. Потапов. – М.: Интеллектуальная литература, 2021.

Степин 2003 – *Степин В.С.* Теоретическое знание. – М.: Прогресс-Традиция, 2003.

Щедровицкий 1975 – *Щедровицкий Г.П.* Автоматизация проектирования и задачи развития проектировочной деятельности // Разработка и внедрение автоматизированных систем в проектирование (теория и методология) / отв. ред. Б.В. Сазонов. – М.: Стройиздат, 1975. С. 9–177.

Espejo 2015 – *Espejo R.* Good social cybernetics is a must in policy processes // Kybernetes. 2015. Vol. 44. No. 6/7. P. 874–890.

Müller 2015 – *Müller K.H.* The Multiple Faces of Reflexive Research Designs // Systemics, Cybernetics and Informatics. 2015. Vol. 13. No. 6. P. 87–98.

Kauffman 2016 – *Kauffman L.H.* Cybernetics, reflexivity and second-order science // Constructivist Foundations. 2016. Vol. 11. No. 3. P. 489–497.

Lefebvre 1986 – *Lefebvre V.A.* Second Order Cybernetics in the Soviet Union and the West // Power, Autonomy, Utopia: New Approaches toward Complex Systems / ed. by R. Trappl. – New York: Plenum Press, 1986. P. 123–131.

Lepskiy 2018 – *Lepskiy V.* Evolution of Cybernetics: Philosophical and Methodological Analysis // Kybernetes. 2018. Vol. 47. No. 2. P. 249–261.

Novikov 2016 – *Novikov D.A.* Cybernetics: from Past to Future. – Heidelberg: Springer, 2016.

Umpleby, Medvedeva, Lepskiy 2019 – *Umpleby S.A., Medvedeva T.A., Lepskiy V.* Recent Developments in Cybernetics, from Cognition to Social Systems // Cybernetics and Systems. 2019. Vol. 50. No. 4. P. 367–382.

REFERENCES

Espejo R. (2015) Good Social Cybernetics is a Must in Policy Processes. *Kybernetes. Vol. 44, n*o. 6/7, pp. 874–890.

Kauffman L.H. (2016) Cybernetics, Reflexivity and Second-Order Science. *Constructivist Foundations.* Vol. 11, no. 3, pp. 489–497.

Lefebvre V.A. (1973) *Conflicting Structures.* Moscow: Sovetskoye radio (in Russian).

Lefebvre V.A. (1986) Second Order Cybernetics in the Soviet Union and the West. In: Trappl R. (Ed.) *Power, Autonomy, Utopia: New Approaches toward Complex Systems* (pp. 123–131). New York: Plenum Press.

Lektorsky V.A. (2001) *Classical and Non-Classical Epistemology.* Moscow: Editorial URSS (in Russian).

Lepskiy V.E. (1998) *The Conception of Subject-Oriented Computerization of Management Activity.* Moscow: RAS Institute of Psychology (in Russian).

Lepskiy V.E. (2010) *Reflexive-Active Environments of Innovative Development.* Moscow: Kogito-Tsentr (in Russian).

Lepskiy V.E. (2016) *Management Technologies in Information Wars (From Classics to Post-Non-Classics).* Moscow: Kogito-Tsentr (in Russian).

Lepskiy V.E. (2017) Reflexive activity in management. *Chelovek i mir.* Vol. 1, no. 1, pp. 53–80 (in Russian).

Lepskiy V. (2018) Evolution of Cybernetics: Philosophical and Methodological Analysis. *Kybernetes.* Vol. 47, no. 2, pp. 249–261.

Lepskiy V.E. (2019) *Methodological and Philosophical Analysis of the Development of Management Problems.* Moscow: Kogito- Tsentr (in Russian).

Lepskiy V.E. (2020) Philosophical and Methodological Grounds for Assessing the Socio-Psychological Consequences of the Introduction of New Technologies. *Psikhologicheskii zhurnal.* Vol. 41. No. 4. P. 105–108.

Müller K.H. (2015) The Multiple Faces of Reflexive Research Designs. *Systemics, Cybernetics and Informatics.* Vol. 13, no. 6, pp. 87–98.

Novikov D.A. (2016) *Cybernetics: From Past to Future.* Heidelberg: Springer.

Rubinstein S.L. (1997) *Selected Philosophical and Psychological Works. Foundations of Ontology, Logic and Psychology.* Moscow: Nauka (in Russian).

Shchedrovitsky G.P. (1975) Design Automation and Development Tasks of Design Activities. In: Sazonov B.V. (Ed.) *Development and Implementation of Automated Systems in Design (Theory and Methodology)* (pp. 9–177). Moscow: Stroyizdat (in Russian).

Stepin V.S. (2003) *Theoretical Knowledge.* Moscow: Progress-Traditsiya (in Russian).

Umpleby S.A., Medvedeva T.A., & Lepskiy V. (2019) Recent Developments in Cybernetics, from Cognition to Social Systems. *Cybernetics and Systems.* Vol. 50, no. 4, pp. 367–382.

Vedyakhin A.A. et al. (2021) *Strong Artificial Intelligence. On the Approaches to the Supermind.* Moscow: Intellektual'naya literatura (in Russian).

DOI: 10.30727/0235-1188-2021-64-1-102-115
Original research paper
Оригинальная исследовательская статья

Architectural Approach to Design of Emotional Intelligent Systems*

A.V. Shiller
Lomonosov Moscow State University, Moscow, Russia

O.E. Petrunya
Moscow Aviation Institute, Moscow, Russia

Abstract

Over the past decades, due to the course towards digitalization of all areas of life, interest in modeling and creating intelligent systems has increased significantly. However, there are now a stagnation in the industry, a lack of attention to analog and bionic approaches as alternatives to digital, numerous speculations on "neuro" issues for commercial and other purposes, and an increase in social and environmental risks. The article provides an overview of the development of artificial intelligence (AI) conceptions toward increasing the human likeness of machines: from the key ideas of A. Turing and J. von Neumann, who initiated the digitalization of society, to discussions about the definition of AI and the emergence of conceptions of strong and weak AI. Special attention is paid to the approach of A. Sloman, to ideas about the architecture and design of complex artificial systems are considered, which make it possible to "emotionally" expand the idea of weak/strong AI. In the article's section on the necessity and possibility of incorporating emotions into the architecture of AI, the authors reveal the goals and methodological limitations for creating an emotional artificial agent. In addition, the article briefly presents the main principles of the authors' architectural approach to the creation of emotional intellectual systems on the example of the cognitive-affective model of architecture, which allow modeling the impact of emotions on the cognitive processes involved in decision-making processes. The described architectural approach to modeling intelligent systems can be used as a conceptual basis for discussing and formulating a strategy for the development of neurocomputing, philosophy of artificial intelligence, and experimental philosophy, for developing innovative research programs, formulating and solving theoretical and methodological problems.

* The study was supported by the Interdisciplinary Scientific and Educational School of Moscow University "Brain, Cognitive Systems, Artificial Intelligence."

Keywords: philosophy of artificial intelligence, weak AI, strong AI, emotions, digitalization, anthropomorphism, intelligent agent.

Alexandra V. Shiller – Ph.D. in Philosophy, Head of the grant's department, Faculty of Philosophy, Lomonosov Moscow State University.
shiller.a@gmail.com
https://orcid.org/0000-0002-2466-9476

Oleg E. Petrunya – Ph.D. in Philosophy, Associate Professor, Department of Philosophy, Moscow Aviation Institute.
hypostasis@yandex.ru
https://orcid.org/0000-0002-5306-5129

For citation: Shiller A.V. & Petrunya O.E. (2021) Architectural Approach to Design of Emotional Intelligent Systems. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 102–115.
DOI: 10.30727/0235-1188-2021-64-1-102-115

Архитектурный подход к созданию эмоциональных интеллектуальных систем*

А.В. Шиллер
Московский государственный университет имени М.В. Ломоносова, Москва, Россия

О.Э. Петруня
Московский авиационный институт (национальный исследовательский университет), Москва, Россия

Аннотация

В течение последних десятилетий из-за курса на цифровизацию всех областей жизни значительно возрос интерес к моделированию и созданию интеллектуальных систем. Однако сейчас наблюдается застой в отрасли, блокирование аналогового и бионического подходов в качестве альтернативных цифровому, многочисленные спекуляции на нейротематике в коммерческих и иных целях, рост социальных и экологических рисков. В статье представлен обзор развития представлений об искусственном интеллекте (ИИ) на пути к повышению человекоподобия машин: от ключевых идей А. Тьюринга

* Исследование выполнено при поддержке Междисциплинарной научно-образовательной школы Московского университета «Мозг, когнитивные системы, искусственный интеллект».

и Дж. фон Неймана, положивших начало цифрофизации общества, до дискуссий об определении ИИ и появлении положений о «сильном» и «слабом» ИИ. Отдельное внимание уделено подходу А. Сломана, рассмотрены идеи об архитектуре и дизайне сложных искусственных систем, благодаря которым становится возможно «эмоциональное» расширение идеи слабого/сильного ИИ. В разделе статьи, посвященном необходимости и возможности включения эмоций в архитектуру ИИ, описаны цели и методологические ограничения для создания эмоционального искусственного агента. Кроме того, в статье кратко представлены основные принципы авторского архитектурного подхода к созданию эмоциональных интеллектуальных систем на примере когнитивно-аффективной модели архитектуры, которая позволяет моделировать влияние эмоций на когнитивные процессы, задействованные в процессах принятия решений. Описанный архитектурный подход к моделированию интеллектуальных систем может быть использован в качестве концептуальной основы при обсуждении и формулировании стратегии развития нейрокомпьютинга, философии искусственного интеллекта и экспериментальной философии, для разработки плодотворных научно-исследовательских программ, постановки и решения теоретических и методологических проблем.

Ключевые слова: философия искусственного интеллекта, слабый ИИ, сильный ИИ, эмоции, цифровизация, антропоморфизм, интеллектуальный агент.

Шиллер Александра Викторовна – кандидат философских наук, руководитель грантового отдела философского факультета МГУ имени М.В. Ломоносова.
https://orcid.org/0000-0002-2466-9476

Петруня Олег Эдуардович – кандидат философских наук, доцент кафедры философии Московского авиационного института (национального исследовательского университета).
hypostasis@yandex.ru
https://orcid.org/0000-0002-5306-5129

Для цитирования: *Шиллер А.В, Петруня О.Э.* (2021) Архитектурный подход к созданию эмоциональных интеллектуальных систем // Философские науки. 2021. Т. 64. № 1. С. 102–115.
DOI: 10.30727/0235-1188-2021-64-1-102-115

Introduction

The relevance of the topic is determined, first of all, by the radical changes that have swept modern society, the “digitalization,” which

is the process of development and introduction of computers and software into the communication and management systems of economy, politics, and society, using the capabilities of high-level programming languages, including their significant independence from specific hardware. The ideological context of digitalization is created by two key ideas: A. Turing's idea about the similarity between a computing machine operatings with a finite number of states and a person with limited memory [Turing 1937] and J. von Neumann's idea about the identity of living organisms and finite automata [von Neumann 1995]. However, the main trend is set by the scenario of the *imitation game* proposed by Turing [Alekseev 2013], if the machine succeeds in this game, it will open the era of *strong artificial intelligence.*

The ultimate goal of modern digitalization is planetary artificial intelligence (AI) and the transformation of human nature itself. This idea is one of the dogmas of neo-mechanicism, a worldview based on a formal numerical understanding of the world. Modern digitalization is the implementation of this worldview in the field of technology, which implies an essential restructuring of the entire scientific and technical methodology in a formal and mathematical way, a finite unification of tools, a rejection of bionic and analog approaches [Petrunya 2017]. The above-mentioned trend has already slowed down the process of creating progressive neurocomputers. Their place was taken by imitation mathematical models that have nothing to do with biological objects.

For this reason, the digital environment was initially perceived with a fundamental strategic flaw – it is unfriendly, non-anthropomorphic, and disproportionate to humans. Therefore, in the real practice of designing computers and creating software for effective work here and now, it is necessary to solve not only many technical problems. There is a need to improve the efficiency of human-machine interaction, first of all, in the development of user and other interfaces that are already actively included in people's lives and are used by banks, telecom operators, and other areas of business [Vedyakhin et al. 2021]. Unfortunately, the solution to these problems occurs spontaneously with all the resulting disadvantages.

Nevertheless, in Western cognitive science, one way or another, the question of increasing the human likeness of machines was raised. First, this is due to efforts to make a machine that is able to pass the Turing imitation criterion. Second, this search has stimulated a set of relevant practical challenges mentioned above, including user-friendliness and usability. Below we will consider two approaches, which, along with

many others, are used in cognitive and computer sciences to increase the human likeness of machines: (a) a design approach in creating computer architectures and (b) modeling emotions in these architectures.

Discussions on the concept of AI and A. Sloman's design approach

The term "artificial intelligence" was coined in the mid-1950s by John McCarthy at the Dartmouth Conferences. A little later, in 1959, McCarthy with Marvin Minsky organized the Computer Science and Artificial Intelligence laboratory at the Massachusetts Institute of Technology. It was McCarthy who became the main ideologist and enthusiast of the field that J. Searle in the article "Minds, Brains, and Programs" called strong AI [Searle 1980]. The concept of strong AI originates in the Turing interpretation of the consequences of the absence of a solution to D. Hilbert's decision problem (*das Entscheidungsproblem*). E. Post provides a different, narrow understanding of the consequences, considering his method and the machine proposed for it as proving "…significant in the development of symbolic logic along the lines of Gödel's theorem on the incompleteness of symbolic logics and Church's result concerning absolutely unsolvable problems" [Post 1936, 103]. A. Turing expanded the problem field in the article "Computing Machinery and Intelligence," proposing to consider the computational actions of a digital machine with a finite number of states as identical to the computational activity of a person with limited (finite) capabilities [Turing 1950]. Thus, D. Hilbert's finitism received a new functionalist interpretation and a new breath. However, it can be noted that, in Turing's problem statement, it is not clear how a negative answer to the question "Can machines think?" may be verified [Yanovskaya 1960]. Such understatement, in our opinion, can be interpreted not so much as a lack of solution to the problem but rather as an uncertainty in the area of its solution.

Another concept, according to which the machine does not think, but imitates thinking, was labeled by J. Searl labeled as weak AI. At the same time, Searl's main intention is emotionally colored; in fact, he wonders how one can agree with McCarthy's thesis about the identity between thermostats, telephones, computing machines, and people? He suggests thinking about the consequences of adopting such beliefs. Against the idea of strong AI, Searle offers the Chinese room argument. The argument describes a situation in which the machine's lack of understanding becomes apparent. Unfortunately, Searle missed

another aspect of the problem: the existing duality of the intellectual activity of the cognizing subject, based on reasoning (discourse) and intuition (speculation) as independent abilities. Although this split dates back to ancient Greek tradition, it was not obvious to all Western intellectuals (including Searle). At the same time, in H. Poincaré's works on the philosophy of mathematics, such a duality is taken into consideration [Poincaré 2014]. Thus, intentionality is not equivalent to reasoning [Biryukov 1978]. It "organizes the subject's attention and his understanding of the object of intellectual or sensory contemplation. Discursive forms are easily formalized, i.e., are replaced by a sequence of signs" [Feinberg 1981, 38], in contrast to the intuitive component of thinking, for which this is impossible.

British researcher A. Sloman replied to criticism of the concept of strong AI with an article with the eloquent title "Did Searle Attack Strong Strong or Weak Strong AI?" [Sloman 1986]. It should be added that Sloman shares the Newell–Simon hypothesis based on the assertion that symbol manipulation is the basis of the ability to think [Newell & Simon 1976] as well as Simon's ideas about the architecture and design of complex artificial systems [Simon 1969]. It is in this vein that one should understand the polemic between Sloman and Searl.

A. Sloman rejects Searl's splitting of AI into strong and weak versions. Based on the assertion that AI accurately reproduces certain human actions that can be characterized in terms of *functions, tasks* and *adaptations* (the Simon–Sloman position), Sloman refuses to accept the thesis that there is a fundamental difference between a human and a "smart" machine. So, he does not even consider a weak version of Searl's AI. Being a supporter of the strong version, Sloman offers its own splitting of AI into weak strong AI and strong strong AI versions. If it cannot be said that AI does not reproduce any human mental operations at all (a weak version of Searl), then there is no need to admit that it reproduces them completely. Marking his own position as weak strong AI and the one that Searl "attacks" as strong strong AI, Sloman states: his opponent is at war with windmills – none of Searl's opponents strongly believes in strong strong AI. But the last thesis hardly reflects the position of J. McCarthy, whom Searle, first of all, criticizes in the above-named article. By the way, McCarthy himself lamented that "as soon as the system starts working normally, they immediately stop calling it artificial intelligence" [Petrunya 2021]. Thus, without denying Sloman the rightness, it should be recognized that his dichotomy does not exhaust all positions in AI, but it complements Searle's analysis.

To further clarify the described controversy, it must be said that in relation to AI, since the late 1950s – early 1960s, Russian philosophy and science has been stuck to the positions that J. Searle marks as weak AI. This context of understanding was formed by S.A. Yanovskaya in her preface to the Russian translation of Turing's article "Can Machines Think?" [Yanovskaya 1960]. However, the Sloman dichotomy can add an important nuance to this analysis: the Soviet scientific community can be divided into supporters of weak weak AI and strong weak AI. In the first case, the researchers did not accept or did not consider any similarity between a machine and a person, in the second, their functional analogy was postulated [Svintsitsky 1964].

In this regard, we should recognize the proximity of Sloman's weak strong version of AI (Simon–Sloman's position) and the Russian version of strong weak AI, which, postulating the similarity of a person and a machine, brackets their differences out. Opening up great opportunities for conceptual speculation under the guise of AI, such an approach simultaneously demonstrates its productivity in the field of putting forward bold hypotheses and their empirical testing. And here Simon-Sloman's conception of the architecture of complexity is quite capable of competing with the P.K. Anokhin's theory of functional systems.

Subsequently, A. Sloman and colleagues [Wright, Sloman, & Beaudoin 1996], generalizing and developing the ideas of Simon, proposed a certain kind of information processing architecture that "emotionally" expands the idea of weak strong AI. This architecture is based not only on the capabilities of high-level programming languages with their significant independence from the hardware implementation of a computer, but mainly on the research methodology of the abstract space of possible requirements for functioning agents (niche space) and the space of possible designs for such agents (design space) and mapping them. Since the architecture dominates over the mechanism in Sloman's approach, the design usually determines the capabilities more than the details of its implementation in each specific case. This, in particular, makes it possible to reconcile the computational and connectionist approaches, i.e., positions of McCarthy and Minsky.

Any architecture is based on the principles envisaged by the designer, and the subsequent functioning and development of the information system does not go beyond those principles. The concept of architecture here is congruent with the systemic characteristics of human thinking

and the psyche as a whole, i.e., causal structures implemented in a non-neural substrate [Wright, Sloman, & Beaudoin 1996].

In the design approach, two interrelated tasks are solved: (1) the research of deeper mechanisms that generate the mental phenomena of living beings; (2) the research of new possibilities for creating perfect intelligent systems. The starting point of the analysis is the research of structures (design) that satisfy the requirements for autonomous self-organizing systems. The design approach allows to successfully explicate the required content into specific design forms and prepare proposals for its subsequent approximation as well as empirical testing of hypotheses, using a variety of interdisciplinary theoretical approaches.

Necessity and possibility of including emotions into AI architecture

Sloman's design approach to the creation of an artificial agent architecture shows the possibility of including in its composition such blocks that were previously proposed to be excluded – the block of motivation, emotions, "morality" (the prototype of the ethical system of an artificial agent). However, Sloman points out that "some AI researchers believe that it should be the goal of AI to design agents that overcome human limitations while displaying all their strengths. This may not be possible if some of the limitations are inevitable consequences of the mechanisms and architectures required to produce those strengths" [Sloman 1999]. That is why it is necessary to understand the phenomenology of emotions, the relationship between emotions and other cognitive functions, methodological limitations and requirements for modeling emotions in the architecture of AI [LeDoux, Phelps, & Alberini 2016; Panksepp 1998].

There are several approaches that are most important for philosophy and cognitive sciences: the three-factor theory of C. Osgood and his colleagues [Osgood, May, & Miron 1975]; the theory of the six components of emotions [Fontaine et. al. 2007]; M. Lewis's theory of the equal status of cognitions and emotions [Lewis 2008]; the theory of embodied cognition [Niedenthal et. al. 2009]; the concept of enactivism [Colombetti 2014]; the theory of cognitive assessment of the emergence of emotions [Ortony, Clore, & Collins 1988]; A. Sloman's concept of architecture of the agent [Sloman 1987]; J. Broekens, D. DeGroot and W. Kosters's approach to modeling emotions [Broekens, DeGroot, & Kosters 2008].

The analysis of these approaches shows that emotions and cognitive processes are functionally connected. Therefore, emotions as an influencing factor that cannot be removed from the system of interrelated processes, should be assigned one of the central places in the architecture of an artificial intellectual system. Simulating emotions in AI becomes especially important when there are such types of tasks of creating AI: (1) achieving system integration by ensuring multimodality of cognitive processes; (2) creation of a flexible architecture that changes during the cycle of functioning of an artificial agent; (3) ensuring high reliability of the interaction of AI with a person; (4) building a digital model of the human emotional system for its theoretical research.

The requirement to include emotions in the architecture of AI is due to both the practical tasks of developing emotional agents and theoretical search that is driven by the "return to corporeality" and the anthropomorphic turn in philosophy and cognitive science. The future-oriented goal is the creation of a model of anthropomorphic intelligence for experimenting with, which will act as an analog of "natural" intelligence.

However, the current possibilities for modeling emotions in artificial agents are extremely limited. Therefore, there should be progress in the transition toward the use of individual theories and modeling parts of the architecture of an artificial agent. In this article, we propose to consider the principles of creating an emotional architecture of an artificial agent using the example of the cognitive-affective model of architecture (CAMA).

Cognitive-affective model of architecture

One of the problems identified in the analysis of the field of emotional modeling is "the lack of a theory that can become the basis for modeling the process of expressing the effects of emotions" [Shiller 2019, 237]. To solve this problem, we propose to investigate the principles of building an architecture model that includes both cognitive and affective components.

CAMA is based on the theory of embodied cognition, the Sloman's concept of architecture and Broekens' approach to modeling emotions. It seems to us that this model, which combines several modern approaches, may turn out to be promising for the purposes of computational modeling of emotions.

The principle of combining cognitive and affective components into one model takes into account and allows modeling individual differ-

ences (personality traits) as well as the impact of emotions (affective states) on cognitive functions involved in decision-making processes (attention, goal-setting, etc.).

Thus, the ***first principle*** of the model is ***synthesis of*** both different components and theoretical foundations.

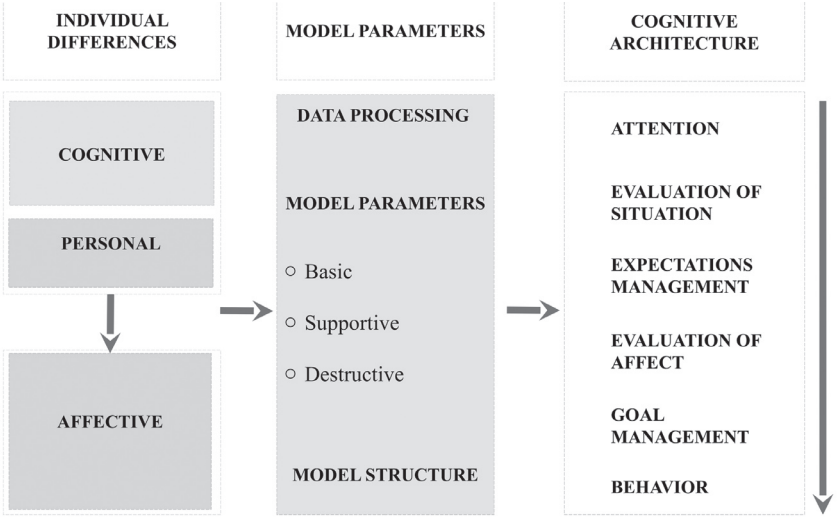

Figure 1. *Diagram of an architecture model*

In the middle part of the diagram (Fig. 1), a set of parameters of the cognitive architecture is presented, among which there should be noted the parameters responsible for data processing (processing speed, working memory), as well as structural parameters (processing rules and regulations, the scheme of intermodular connections and their weight).

The work of the model can be represented in several steps:

(1) Input: obtaining perceptual data (from the external environment), which may be important for making decisions and choosing actions, as well as the presence of predetermined information about the characteristics of the agent (a certain "synthetic personality," for which there are envisaged parameters of the intensity of emotions, the value of various personality traits).

(2) Data processing: it depends on individual (cognitive and affective) differences and occurs by correlating the predetermined characteristics with the architecture parameters, which determines changes in the course of data processing (acceleration, deceleration, reduction in the set

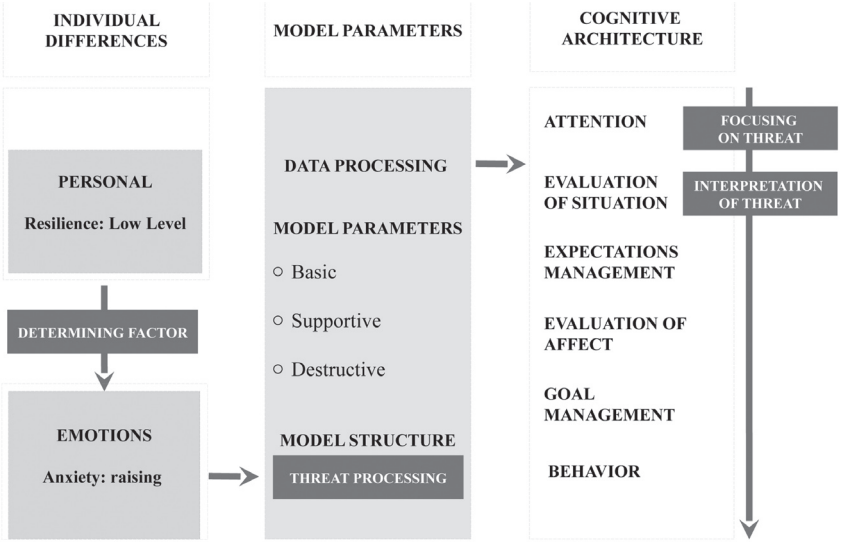

Figure 2. *Impact of low resilience and anxiety on threat perception*

of steps, prioritization of the processing of certain data). The matching mechanism is based on empirical data and ideas about how a person acts with a similar set of perceptual data and personality characteristics.

(3) Output: there is a hierarchical system containing a set of situations and goals in which the initial perceptual data and the proposed behavior are differently correlated. Data analysis allows us to assign some "tags" to information – for example, if the received perceptual data is assessed as dangerous, posing a threat to the life of the agent, then the role of the constructs of anxiety and fear increases, which determines the speed of data processing in the modules of attention and situation assessment as well as the choice of the agent's behavior as a result of data analysis (see Fig. 2).

Over time, similar perceptual data will be more quickly tagged as dangerous and, therefore, prioritized for processing.

As Fellous points out, "emotions are patterns of neuromodulations that affect brain areas involved at all levels of functions, from low-level motor control to planning and high-level cognition" [Fellous 2004, 7]. Therefore, to express the cognitive effects of emotion the model uses parametric functions consisting of linear sets of factors that are estimated and affect each of the parameters, which is consistent with the data of modern research. An example is the following parameter: "working memory capacity reflects the normalized weighted sum of emotion intensities, trait values, and skill level" [Shiller 2019, 236].

The second principle of the model is ***anthropomorphism,*** that is, the human psyche is the basis of the architecture of the model and the dominant that determines the emotionality and behavior of the artificial agent. Based on these two principles, the features of the model are as follows:

(1) The ability to create many variants of synthetic personalities by defining a set of cognitive and affective components, personality traits.

(2) The consideration of the role of corporeality and, as a consequence, the ability to implement the principles of embodied cognition in creating AI.

(3) The ability to partially implement the principle of neuroplasticity, which underlies the work of the human brain.

(4) The presence of restrictions associated with the need to formalize the processes of the human psyche, which inevitably leads to the simplification of the final model of architecture.

In the future, it is planned to expand and improve this architecture model by testing primary hypotheses, collecting data and implementing theoretical provisions in order to form an acting artificial agent with an emotional subsystem based on CAMA.

Conclusion

Nowadays, there is a huge rise in scientific and pseudo-scientific research on AI and its mention in the media. However, it should be noted that, in most cases, all these researches are of applied nature and use the phrase "artificial intelligence" as a catchy analogue of fairly simple technologies that can only be attributed to the "weak" version of AI. The reason lies in the fact that the digital approach to the development of AI technologies for a long time was understood as the only "correct" one, and analog and body-oriented approaches, which consider the multimodality of cognitive activity and possible errors/imperfections that ensure the richness of this activity, were excluded from the sphere of interest of researchers.

The described discussions on the definition and various versions (strong, weak, or some intermediate one) of AI illustrate the complexity of both the concept and the scientific direction. As an inevitable consequence of the development of this research area, there are an anthropomorphic turn in the studies on AI and efforts to model analogs of human intelligence or at least some of its constituent parts. Sloman's model of architecture for an artificial agent, Broekens's researches on modeling,

and numerous experimental data in line with the theory of embodied cognition inspired the authors of the article to create another version of cognitive architecture that allows modeling individual differences and affective states in the process of cognitive data processing. In the future, additional studies are planned that will allow testing hypotheses, supplementing the architecture with new blocks, forming a mathematical apparatus, and moving from theory to practice, implementing the model in the form of a digital assistant/bot that is capable of the believable expression of emotion.

REFERENCES

Alekseev A.Yu. (2013) *Complex Turing Test: Philosophical-Methodological and Socio-Cultural Aspects.* Moscow: IInteLL (in Russian).

Biryukov B.V. (1978) What Can Computing Machines Do? Instead of an Afterword. In: Dreyfus H. *What Computers Can't Do. Critique of Artificial Intelligence* (pp. 298–332). Moscow: Progress (in Russian).

Broekens J., DeGroot D., & Kosters W.A. (2008) Formal Models of Appraisal: Theory, Specification, and Computational Model. *Cognitive Systems Research.* Vol. 9, no. 3, pp.173–197.

Colombetti G. (2014) Ideas for an Affective Neuro-physio-phenomenology. In: Colombetti G. *The Feeling Body: Affective Science Meets the Enactive Mind* (pp. 135–170). Cambridge, MA: MIT Press.

Feinberg E.L. (1981) *Cybernetics, Logic, Art.* Moscow: Radio i svyaz' (in Russian).

Fellous J.M. (2004) From Human Emotions to Robot Emotions. In: *Proceedings of the AAAI Spring Symposium: Architecture for Modeling Emotion* (technical report SS-04-02). Menlo Park, CA: AAAI Press.

Fontaine J.R.J., Scherer K.R., Roesch E.B., & Ellsworth P.C. (2007) The World of Emotions Is Not Two-Dimensional. *Psychological Science.* Vol. 18, no. 2, pp. 1050–1057.

LeDoux J., Phelps L., & Alberini C. (2016) What We Talk about When We Talk about Emotions. *Cell.* Vol. 167, no. 6, pp. 1443–1445.

Lewis M. (2008) Self-Conscious Emotions: Embarrassment, Pride, Shame, and Guilt. In: Lewis M. & Haviland-Jones J.M. (Eds.) *Handbook of Emotions* (3rd ed., pp.742–749). New York: Gilford Press.

Newell A. & Simon H.A. (1976) Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM,* Vol. 19, no 3, pp. 113–126.

Niedenthal P.M., Winkielman P., Mondillon L., & Vermeulen N. (2009) Embodiment of Emotion Concepts. *Journal of Personality and Social Psychology.* Vol. 96, no. 6, pp.1120–1136.

Ortony A., Clore G.L., & Collins A. (1988) *The Cognitive Structure of Emotions.* Cambridge, UK: Cambridge University Press.

Osgood C.E., May W.H., & Miron M.S. (1975) *Cross-Cultural Universals in Affective Meaning.* Urbana: University of Illinois Press.

Panksepp J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions.* New York: Oxford University Press.

Petrunya O.E. (2017) The Opposition of Natural and Artificial in Neuroscience: "Who Is to Blame" and "What to Do." *Neyrokompyutery: razrabotka, primeneniye = Neurocomputers: Development, Application.* No. 4, pp. 26-30 (in Russian).

Petrunya O.E. (2021) *Anthropic Approach vs. Imitation Game: Philosophical and Methodological Problems of Science and Technology in the Digital Age. M*oscow: LENAND (in Russian).

Poincaré H. (2014) Mathematics and Logic. In: *The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method* (pp. 448–459). Cambridge, UK: Cambridge University Press.

Post E.L. (1936) Finite Combinatory Processes – Formulation 1. *The Journal of Symbolic Logic.* Vol. 1, no. 3, pp. 103–105.

Searle J.R. (1980) Minds, Brains, and Programs. *Behavioral and Brain Sciences*. Vol. 3, no 3, pp. 417–457.

Shiller A.V. (2019) Expression of Modeled Effects of Emotions in Artificial Agents as a Visual Language. *ΠΡΑΞΗΜΑ. Problems of Visual Semiotics.* Vol. 22, no. 4., pp. 223–243 (in Russian).

Simon H.A. (1969) *The Sciences of the Artificial.* Cambridge, MA: MIT Press.

Sloman A. (1986) Did Searle Attack Strong Strong or Weak Strong AI? In: Cohn A.G. & Thomas J.R. (Eds.) *Artificial Intelligence and Its Applications* (pp. 271–288). Chichester: John Wiley and Sons.

Sloman A. (1999) What Sort of Architecture is Required for a Human-Like Agent? In: Wooldridge M. & Rao A. (Eds.) *Foundations of Rational Agency* (pp. 35–52). Dordrecht: Springer.

Svintsitsky V.N. (1964) On the Question of the Genetic Connection of Cybernetics with Classical Automatics. In: Berg A.I., Biryukov B.B., Novik I.B., Kuznetsov I.V., & Spirkin A.G. (Eds.) *Cybernetics, Thought, Life* (pp. 164–172). Moscow: Mysl' (in Russian).

Turing A. (1937) On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society. Series 2.* Vol. 42, no. 1, pp. 230–265.

Turing A. (1950) Computing Machinery and Intelligence. *Mind.* Vol. 59. no. 236, pp. 433–460.

Vedyakhin A.A. et al. (2021) *Strong Artificial Intelligence: Approaching the Supermind* (A.S. Potapov, Ed.). Moscow: Alpina Publisher (in Russian).

von Neumann J. (1995) The General and Logical Theory of Automata. In: Bródy F. & *Vámos T. (Eds.) The Neumann Compendium* (Vol. 1, pp. 526–566). Singapore: World Scientific.

Wright I.P., Sloman A., Beaudoin L.P. (1996) Towards a Design-Based Analysis of Emotional Episodes. *Philosophy, Psychiatry, & Psychology.* Vol. 3, no 2, pp.101–126.

Yanovskaya S.A. (1960) Preface to the Russian Translation. In: Turing A. *Can Machines Think?* (pp. 3–11). Moscow: Fizmatgiz (in Russian).

КОГНИТИВНОЕ ПРОСТРАНСТВО

Электронная культура: проблемы и перспективы

DOI: 10.30727/0235-1188-2021-64-1-116-133
Оригинальная исследовательская статья
Original research paper

Наука и искусство в цифровую эпоху: проблема синергии

В.Г. Буданов
Институт философии РАН, Москва, Россия

А.Р. Ефимов
ПАО «Сбербанк», Москва, Россия
Национальный исследовательский технологический университет «МИСиС», Москва, Россия

Аннотация

В статье речь идет о том, что сегодня художественное восприятие мира и научно-техническое понимание реальности остаются основными формами творческой самореализации людей. Согласно авторской позиции, на протяжении многих веков, начиная с Античности, искусство и наука шли рука об руку в культуре. Однако в период расцвета техногенной цивилизации происходит раскол, и эти две части ранее единой культуры со второй половины XX века становятся плохо совместимыми. По утверждению авторов, эпоха цифровизации окончательно заменит алгоритмизируемые и инструктивные типы профессий техников, обслуживающего персонала робототехникой и искусственным интеллектом, а человек вынужден будет развивать себя в сферах правополушарных практик, к которым, несомненно, относятся научно-техническое творчество и искусство. Как и все предыдущие информационные революции, современная цифровая революция создает новые сетевые структуры быстрой коммуникации и сверхдальнего, глобального порядка. Расстояние при этом не играет роли. В данной ситуации радикально изменяются когнитивные карты человека, возникают иные типы самоорганизации и социализации, происходит деформация ценностных пространств и мировоззренческих ориентиров. Жизнь в неопределенности нового мира может стать эффектив-

ной только во взаимодействии человека и искусственного интеллекта, и единственным аспектом деятельности, который нельзя заменить искусственным интеллектом, являются творчество и эстетические переживания, культура. С учетом этого в заключение авторами сделан вывод о том, что в настоящее время в нам необходим новый синтез в культуре. Пути такого синтеза подробно рассмотрены в статье. Среди них – продвижение новых концепций *Science-Art*, применение постньюринговой методологии, использование квантово-синергетической антропологии, развивающей новые представления о театральном и инженерно-техническом творчестве, недоступном для искусственного интеллекта в ближайшем будущем.

Ключевые слова: проблема двух культур, цифровизация, техногенная цивилизация, постньюринговая методология, искусственный интеллект, творчество, искусство, наука.

Буданов Владимир Григорьевич – доктор философских наук, кандидат физико-математических наук, главный научный сотрудник, руководитель сектора междисциплинарных проблем научно-технического развития Института философии РАН.

budsyn@yandex.ru

http://orcid.org/0000-0003-2371-8659

Ефимов Альберт Рувимович – вице-президент, директор управления исследований и инноваций ПАО «Сбербанк», заведующий кафедрой инженерной кибернетики Национального исследовательского технологического университета «МИСиС».

makkawity@gmail.com

https://orcid.org/0000-0001-6857-8659

Для цитирования: *Буданов В.Г., Ефимов А.Р.* Наука и искусство в цифровую эпоху: проблема синергии // Философские науки. 2021. Т. 64. № 1. С. 116–133. DOI: 10.30727/0235-1188-2021-64-1-116-133

Science and Art in the Digital Age: A Synergy Problem

V.G. Budanov
Institution of Philosophy, Russian Academy of Sciences, Moscow, Russia

A.R. Efimov
PJSC Sberbank, Moscow, Russia;
National Technology Research University MISiS, Moscow, Russia

Abstract

Today artistic perception of the world and scientific and technical understanding of reality remain the main forms of creative self-realization. For many centuries, starting in Antiquity, art and science went hand in hand in cultural history. However, during the heyday of technogenic civilization, there occurred a split, and since the second half of the 20th century these two parts of a previously common culture became poorly compatible. According to the authors, the era of digitalization is going to completely replace the algorithmic and instructive professions in technology and service personnel with robots and artificial intelligence, and a person will have to develop the spheres of right-brain practices, which undoubtedly include scientific and technical creativity and art. Like all previous information revolutions, the modern digital revolution is creating new network structures of fast communication and ultra-long-distance, global order. Distance does not matter any more. In this situation, the cognitive maps of a person change radically, new types of self-organization and socialization appear, and there occurs a deformation of value spaces and worldview guidelines. Life in the uncertainty of this new world can only become effective in interaction of man and artificial intelligence, and the only aspect of activity that cannot be replaced by artificial intelligence is creativity and aesthetic experiences as well as culture. Thus, the authors conclude that a new synthesis in culture is needed, the ways of which are discussed in detail in the article. Among them are promotion of new concepts of Science-Art, application of post-Turing methodology, and use of quantum-synergetic anthropology, which develops new ideas about theatrical and engineering creativity, – and these will remain unattainable to artificial intelligence in foreseeable future.

Keywords: problem of two cultures, digitalization, technogenic civilization, post-Turing methodology, artificial intelligence, creativity, art, science.

Vladimir G. Budanov – D.Sc. in Philosophy, Ph.D. in Physics, Chief Research Fellow, Head of the Department of Interdisciplinary Problems in the Advance of Science and Technology, Institute of Philosophy, Russian Academy of Sciences.

bvg55@yandex.ru

https://orcid.org/0000-0003-2371-8659

Albert R. Efimov – Vice-President, Director of Research and Innovation, PJSC Sberbank; Head of Department of the Engineering Cybernetics, National University of Science and Technology MISiS.

makkawity@gmail.com

https://orcid.org/0000-0001-6857-8659

For citation: Budanov V.G. & Efimov A.R. (2021) Science and Art in the Digital Age: A Synergy Problem. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 116–133.
DOI: 10.30727/0235-1188-2021-64-1-116-133

Введение

Наука и искусство, наряду с философией, религией, мифологией и обыденным сознанием, порождают различные формы познания и понимания мира современного человека [Степин 2000]. Если религиозная и мифологическая картина мира угасает в настоящее время, а обыденная форма сознания все больше технологизируется и алгоритмизируется, напоминая обрядовый магизм древних (рецептурная техномагия и технофетишизм), то именно художественное восприятие мира и научно-техническое понимание реальности остаются основными формами творческой самореализации людей.

Исторически искусство рождается первым. Вместе с культами первобытного человека оно тысячелетиями было пронизано мифом и религиозным содержанием. Философия, наука возникают последними в античной Греции, а в современном понимании – лишь в Новое время. Подчеркнем, что эти формы имманентны человеку изначально. Каждый ребенок, независимо от эпохи, проходит стадию мифологического и магического ощущения мира, закрепляемого сказками, становится художником и мастером еще в дошкольном возрасте. Представление о науке и технике формируется у него уже в школе. Но философия открывается ему только с жизненным опытом.

Место эстетического в познании

Иммануил Кант писал о том, что две вещи непрестанно вызывают его восхищение: «звездное небо надо мной и нравственный закон во мне». Если нравственный закон воплощен в этических основаниях мифов и религий, то восхищение бездной звезд и тайной мироздания характерны для искусства и науки, которые всегда находились рядом в античной Греции. Красота была критерием научности более двух тысяч лет, со времен школы Пифагора и до откровений Кеплера, предложившего свои законы движения планет из соображения красоты, а понятие «техне» трактовали как искусство художника и поэта, как мастерство ремесленника и инженера.

Примеры такого синтеза науки и искусства существуют не только в период античной и средневековой Европы, но и в период арабского Возрождения, когда научные трактаты писали в стихах. Явление синтеза наблюдается и в эпоху европейского Возрождения в гениальных инженерных прозрениях Леонардо, позднее – в созданном Гёте непревзойденном трактате о законах цветового зрения, который он ставил выше своих поэтических шедевров. Даже в конце XIX века не воспринималось как парадоксальное то обстоятельство, что наш химик А.П. Бородин был и великим композитором, автором всемирно известной оперы «Князь Игорь». Ребенком П.И. Чайковский приобщался к классической музыке в маленьком городке Воткинске, на домашних концертах отца, горного инженера, и его коллег, которые разучивали последние музыкальные новинки, поскольку, согласно уставу горной службы того времени, им предписывалось на досуге музицировать, чтобы культурно не деградировать в российской глубинке. Идеалы воспитания целостной личности в классическом образовании конца XIX – начала XX века все еще доминировали, и на галерке Большого театра в равной степени можно было встретить студентов Московского университета, изучающих гуманитарные дисциплины, и воспитанников Императорского Московского технического училища, осваивающих технические дисциплины.

Разрыв между естественниками, технарями и гуманитариями увеличивается к середине XX века, после двух мировых войн, после развернувшейся напряженной технологической гонки вооружений. Это связано и с нарастающей потребностью достижения высокого профессионализма, личного успеха в максимально короткие сроки, когда образовательные стратегии направлены на подготовку мастерства в узком профиле, а целостное мировоззрение фактически подменяется профессиональной картиной мира. Техногенная цивилизация, в соответствии с терминологией В.С. Степина [Степин, Кузнецова 1994], все чаще использует технологическое мышление, основанное преимущественно на левополушарном, инструктивном стиле мышления, воспроизводятся армии конвейерных рабочих и техников, обслуживающих машины и оборудование.

Творчество инженера и ученого начинают противопоставлять гуманитарному, подчеркивая их различия в отношении предмета и метода. В 1959 году британский физик, химик, писатель-реалист Ч. Сноу заявляет о концепции двух культур в своей книге под на-

званием «Две культуры и научная революция», которая становится резонансной и провоцирует множественные дискуссии в среде мировой интеллигенции [Сноу 1973]. Сноу имел право говорить от лица обеих культур – естественнонаучной и гуманитарной, поскольку был признанным авторитетом этих традиций.

В нашей стране они известны как дискуссии «физиков» и «лириков» 60-х годов, в которых участвовали ведущие поэты и писатели, ученые и инженеры. Оказалось, что многие ученые, имеющие глубокие результаты, активно интересовались, иногда и на профессиональном уровне, вопросами, далекими от наук о природе и технике, проблемами культуры и искусства, истории и психологии, были живописцами, исполнителями и поэтами, т.е. людьми единой культуры в идеалах эпохи Возрождения. В частности, вспомним об одном из основателей отечественной космонавтики, соратнике С.П. Королева, физике-механике, академике Б.В. Раушенбахе, который одновременно был председателем Научного совета Российской академии наук по комплексной проблеме «История мировой культуры», автором фундаментальных трудов по истории России, богословию и теории перспективы в изобразительном искусстве [Балашова 2018; Батурин, Крючков 2017]. Традиции синтетического творчества ученых продолжают и наши современники. Один из самых молодых нобелевских лауреатов по физике, создатель графена, выпускник Московского физико-технического института К.С. Новосёлов является не только покровителем современного искусства, но и профессиональным дизайнером и каллиграфом[1].

Исследователями установлено, что правополушарная активность мозга, связанная с воображением, синтезом и интуицией, которая характерна для художественного творчества, не в меньшей степени отвечает и творчеству математиков, естественников, инженеров. Она происходит не в непрерывном режиме, а перемежается периодами активности логического оформления и проверки интуитивной догадки, практической реализации в материале и испытания в эксперименте, когда более активно левое полушарие. Отличие физиков от лириков состоит не в абсолютном доминировании одного из способов мышления, а в различной форме их микширования. Перефразируя Сальери, можно предположить, что «физики» поверяют алгеброй гармонию чаще (это требова-

[1] С художественными работами К.С. Новоселова можно ознакомиться на его сайте (https://www.kostyanovoselov.com/paintings).

ние научности), чем «лирики». Последние иногда предоставляют проверку другим. «Эврика» в науке аналогична и прозрениям в искусстве, и духовным озарениям, научные инсайты подобны сатори, как утверждает мастер и популяризатор дзен-буддизма Д.Т. Судзуки.

Таким образом, высокая наука, техника и искусство остаются во взаимопонимании и взаимном интересе, а проблема двух культур существует в массовом сознании, образовании и технократическом шовинизме огромного сонма техников, операторов, обслуживающего персонала, живущего по инструкциям (не будем путать с инженерами и технологами). Попытки ее решения предпринимаются на протяжении тридцати лет через введение курсов естествознания в программы гуманитариев, а естественников пытаются приобщать к истории мировой культуры [Буданов 1994]. Это полезно, но радикально проблему не решает, на наш взгляд.

Полагаем, можно надеяться на преодоление раскола культуры в эпоху грядущей цифровизации, как это не покажется странным, и именно благодаря процессам цифровизации. Подобно тому, как кибернетика и автоматизация поточных линий освободила от рутинного алгоритмического труда миллионы конвейерных рабочих, так и введение роботизации и искусственного интеллекта (ИИ) освободит огромное количество техников, операторов и служащих от рутины жестко инструктивных обязанностей для более творческих форм жизни. Конечно, это становится неким экзистенциальным досуговым вызовом для миллионов работающих в богатых странах и драмой безработицы – в бедных. Решение данной проблемы совершенно не тривиально. Попытаемся построить путь к нему.

Единственным преимуществом человека перед машиной остается возможность творчества в самых разных формах. Следовательно, необходимо вернуть творческое начало в максимальное число обыденных и профессиональных практик, найти красоту и креатив в повседневности и технических профессиях, видеть и создавать красоту в быту и работе, стать дизайнером своей среды обитания, производства и его результатов. Подобное отношение к жизни находит отражение в калокагатии греков и учении китайцев – фэншуй. Цифровая эпоха потребует, по нашему мнению, такого отношения от каждого, кто хочет остаться человеком. Искусство становится образом жизни, но и ее научно-технический аспект усиливается многократно. Мы будем жить преимущественно

в цифровых техносредах, а следовательно, процесс их создания и функционирования должен быть сопряжен с эстетическими состояниями человека. Тем самым возникает вызов союза, а не просто дополнения, творчества технического и художественного. В этом, на наш взгляд, состоит главная задача *Science-Art*, т.е. развитие наук о человеке и художественного творчества на фоне бурного развития техносферы. Новая техноантропосфера, вопреки всеобщим опасениям, вполне способна развивать целостную творческую личность.

Приведем и более убедительное суждение: развитие целостной творческой личности есть условие устойчивого развития техноантропосферы. Действительно, сегодня в качестве такого условия выступает концепция техно-гуманитарного баланса или баланса физических и гуманитарных технологий. Задача последних – согласовать изменения в «быстром» мире техносферы с изменениями в «медленном» мире социосферы. Согласование обеспечивается гуманитарными технологиями контроля и управления, которые инсталлируются через программы, схожие с программами цифровой экономики в России. Однако развитие и физических, и гуманитарных технологий опирается на информацию, отчуждаемую от человека. Последовательное развитие в данном направлении приводит к выведению человека «за скобки» этого развития. Баланс может быть обеспечен добавлением в схему техно-гуманитарных технологий, технологий работы со смыслами или неотчуждаемой от человека информацией [Вольнов 2018]. Треугольник физических, гуманитарных и семантических технологий позволяет накопить социальную энергию, вырабатываемую в расколе современной культуры на науку и искусство, предусмотреть конструктивное направление ее на обеспечение устойчивого развития. Одно из разрабатываемых направлений – технология *Science-Art* – семантическая технология работы через мышление с семантическим вакуумом (бесконечностью), где ориентация сознания в бесконечном пространстве смыслов обеспечивается обращением к Красоте [Вольнов 2019].

Модусы научной и творческой деятельности

Известное утверждение о том, что красота спасет мир, может стать не метафорой, а руководством к действию. Попытаемся установить, каким образом содержательно и по форме могут взаимодействовать наука и искусство. Для этих целей следует

вспомнить о том, что наука и искусство относятся к определенным видам деятельности. Процесс каждого из указанных видов деятельности можно условно разделить на четыре причинно-следственных этапа-компонента [Щедровицкий 1997]:

$$S – T – O = R.$$

Соответствующие обозначения представляют собой аббревиатуры деятельностной тетрады: *S – Subject*, *T – Tools*, *O – Object*, *R – Result*. Далее звездочкой (*) обозначим элементы деятельностной тетрады искусства. Элементы научной тетрады будем указывать без звездочки.

При рассмотрении одного деятельностного процесса с позиций другого, т.е. если один из них является объектом другого (например, требуется оценить с позиции человека искусства научную деятельность), получаем следующее представление описания комплексной деятельности:

$$S^* – T^* – (S – T – O = R) = R^*.$$

Или более кратко и символично в виде *S*T**(*STOR*) *R**, т.е. речь идет об оценке конкретной научной деятельности с позиций определенной деятельности (ее образов и методов) из области искусства. Каждый вид деятельности может осуществляться в одном из двух модусов:

Ex – Explore – наблюдательно-исследовательский, где цель заключается в том, чтобы понять и раскрыть сущностные свойства объекта. Данная форма деятельности ближе к процессам в математике, теоретических дисциплинах и процессах проведения научного эксперимента на завершающей стадии, поэтическим и художественным репрезентациям в искусстве;

Tr – Transform – проектно-преобразовательный, где цель состоит в изменении свойств объекта с учетом замысла. Такая форма ближе к литературным, прикладным и исполнительским формам искусства, процессам постановки эксперимента и техническому, инженерному творчеству.

Приведем пример. *Ex* (*S*T*O*R**) – это фото или классический этюд на пленэре, минимально искажающий реальность. *Tr* (*S*T*O*R**) – построение инсталляции или мизансцены природы в театре, дающее большой простор макетирования в материале. В отношении науки *Ex* (*STOR*) – наблюдение солнечной активности на освоенном телескопе; *Tr* (*STOR*) – построение и отлаживание работы этого телескопа или Большого адронного коллайдера в Европейской организации по ядерным исследованиям (ЦЕРН).

В контексте рассмотрения вопроса о применении научного подхода в искусстве ситуация *Ex* (*ST* (*S*T*O*R**) *R*) означает усмотрение научных принципов, свойств и закономерностей в существующей практике искусства, например, акустических феноменов при исполнении музыки в определенном помещении. Речь идет, скорее, о мониторинге. Активное научное творчество в искусстве *Tr* (*ST* (*S*T*O*R**) *R*) в аспекте приведенного выше примера будет означать, что наука (музыкальная акустика) предложит изменения в интерьере, меры по улучшению качества звучания в этом помещении.

Наука и техника за период более ста последних лет радикально изменила инструментарий и объект создания произведения искусства *T**и *O**, способы его донесения до потребителя. Помимо классических кисти и резца возникли метод фотографии в различных участках спектра, кинематограф, телевидение, электронные музыкальные инструменты и способы оркестровки, новые материалы и краски, сложнейшие архитектурные сооружения, безграничные расширения аудитории за счет трансляции и записи исполнения произведений, перформансов и т.д. Этот процесс продолжается и сегодня на базе цифровых технологий с применением голографической пространственной светописи, созданием цифровых виртуальных миров и дополненной реальности.

Наблюдатель от искусства также может быть представлен в двух состояниях – пассивном и активном. Пассивный модус *Ex* (*S*T**(*STOR*) *R**), т.е. человек ищет эстетическое и художественное содержание в процессе реальной научной деятельности. В частности, вспомним о том, насколько романтично и красиво это изображено в фильме «Девять дней одного года», где исполнителем главной роли физика-ядерщика стал А. Баталов. В том числе и подобные фильмы в СССР вызвали мощную волну интереса к физике. Популярность последней создала огромные конкурсы на физические факультеты в конце 60-х годов.

Существует и иная сторона пассивной позиции. Например, сегодня в школьных программах Японии введен обязательный предмет «Любование природой», который, безусловно, развивает эстетические чувства учащихся. Но наука открывает человеку новые горизонты природного, лежащие вне доступа наших органов чувств, переводя нас в другие масштабы пространства, времени, частоты, энергии. Самое простое – это супермакро в фотографии, когда миры крошечных существ, насекомых, гусениц превра-

щаются в миры монстров для фантастических сериалов (такие образы и вдохновляют художников блокбастеров). Погружаясь в наномасштабы новых материалов с помощью электронных микроскопов или разглядывая далекие галактики в супертелескопы, мы зачарованы богатством палитры и грандиозностью ландшафтов загадочных миров. Увеличивая частоты обращения планет по октавному принципу или частоты ионосферы, можем «услышать» пифагорову «музыку сфер» Солнечной системы [Буданов 2017, 282], многократно уменьшая частоты спектров излучения космоса или макротел, способны «услышать» шумы жизни их микромира [Ballora 2014; Quinton, McGregor, Benyon 2016].

Интересен эвристический способ переносов принципов и механизмов природных объектов в технику, который находится в основе бионики и который рождается из внимательного мониторинга природы в самых разных ее масштабах. При этом происходит поиск не столько красоты, сколько подсказки для новой технической идеи. Наша вторая природа – техника – может доставлять не меньшее эстетическое удовольствие при должном отношении к ней. Глядя на раскладку проводов электрооборудования космического корабля, адронного коллайдера[2], суперкомпьютера, нейросети или магистрального самолета, любой из нас способен увидеть грандиозную рукотворную вселенную сложности.

Итак, толкование понятия «любование природой» в широком смысле будет способствовать популяризации и привлекательности науки и техники, а цифровые технологии позволят в большей степени картографировать бесконечные ландшафты в базы *BigData*, чтобы осмысленно работать с ними методами ИИ.

Активный модус – *Tr* (*S*T**(*STOR*) *R**) – это не просто поиск художественного в науке, но и соучастие в дизайне, эстетическом наполнении научной деятельности и научной среды работы ученых. Соответствующая форма сотрудничества науки и искусства еще должна активно развиваться. Пока она редко проявляется в привлечении архитекторов-концептуалистов, художников для оформления зданий и интерьеров научных центров. Однако речь идет о более масштабном сотрудничестве в области *Science-Art* и научно-технического творчества, что должно стимулировать работу инженеров и ученых, привлекать молодежь в науку, создавать ее футурологический образ.

[2] См. работы, посвященные эстетике Большого адронного коллайдера: [Ginter, Heuer, Franzobel 2011; Schwärzler 2016, 137–150].

Помимо четырех описанных форм взаимодействия науки и искусства возможны и гибридные их композиции. Например, если субъектом произведения искусства становится природный феномен, объект техники или активность живого существа, т.е. объект науки *O* становится до некоторой степени субъектом искусства *S**, а инструментом искусства служат методы презентации процессов, осуществляемых этим объектом, без участия человека. Так, красивый рисунок облаков на закатном небосводе, искривленные ветром кроны деревьев на вершине морского утеса зачастую восхищают и наверняка не вызывают возражений у зрителей. И то, и другое создано стихиями природы, почитаемо в любой культуре. Вместе с тем следы борьбы животных на песке, звуки перемещения мухи на мембране микрофона, следы подкрашенного таракана, оставленные на листе бумаги, обычно вызывают недоумение. Требуется мощный концептуальный мотив, чтобы начать искать в этом позитивное эстетическое переживание. Следовательно, экспериментальное искусство уместно и возможно, но не ради эпатажного порыва, а для дальнейшей дискуссии о символических основаниях культуры и нашего восприятия.

Антропоморфизм техники

Многочисленные размышления и обсуждения вызывают сегодня эстетические интерпретации реальных перспектив жизни людей в цифровую эпоху: соединения телесности человека и современных экзоскелетов, перспектив киборгизации, новой урбанистики, интернета вещей, гибридных человеко-машинных систем, антропоморфных роботов, аватаров и т.д. Формируется широкое поле творчества не только для сотрудничества художников и технодизайнеров, философов и антропологов, но и когнитивистов, специалистов по ИИ и робототехнике. Особенно интригующим является проект развития робототехники в рамках посттьюринговой методологии, в котором взаимопонимание с ИИ в немалой степени будет зависеть от общих ценностей и стилей мышления человека и машины [Efimov 2020], включающих в себя и эстетические предпочтения. Ключом к построению обобщенной телесности будущих аватаров могут стать языки, рожденные квантово-синергетической антропологией [Буданов 2012], как для человека, так и для антропоморфных роботов.

Современная наука создает искусственного человека по образу и подобию живого. Для создания антропоморфных машин,

включенных в человеческий социум, требуется эмоциональный искусственный интеллект [Schuller, Schuller 2018; Yonck 2020; Chursinova, Stebelska 2021]. Это дает возможность создавать антропоморфных роботов, способных к имитационному эмоциональному диалогу, что находит спрос и перспективу в практиках воспитания, образования и гендерных отношений, социальных тренингов, вплоть до создания цифровой «копии» личности.

В основе создания антропоморфных нейросетей ИИ, в том числе эмоциональных, находятся процессы глубокого обучения (*deep learning*) на базах данных человеческого поведения и опыта. Освоение машинами обобщенной телесности человека – приоритетная задача создания таких сетей. Интегральная онтология обобщенной телесности могла бы стать ключевым структурно-функциональным способом таксономии человеческой феноменологии телесности. Базой больших данных могла бы послужить запись (видеография) театральных экспликаций тех или иных состояний тел и подтел интегральной онтологии человека в проекции на разные формы и сценарии театрального действия. Таким образом, необходимо предварительное создание «нотации и партитуры» театрального действия на языке интегральной онтологии. Можно обучать нейросеть ИИ подобно тому, как обучают актеров. Именно для этого мы создаем язык, отделимый от личности режиссера-мастера, т.к. многое мастер дает ученикам через эмпатию, а у машины этот канал восприятия отсутствует [Буданов, Синицына 2020].

Обучающие множества будущих нейросетей должны быть подобраны и с позиций человеческого понимания красоты и этики. Рассматривать и пытаться решать эти проблемы логично уже сегодня. В противном случае аватары с примитивным пониманием указанных вопросов начнут обучать следующие поколения наших детей.

Новые формы искусства – *Science-Art*

В качестве примера синтеза науки и искусства приведем основные проблемы и подходы к их решению для образования учащихся естественнонаучных, технических и математических профилей с привлечением различных форм искусства – *STEAM* (Science, Technology, Engineering, Arts, Mathematics)[3].

[3] Этой теме был посвящен круглый стол, организованный в ноябре 2020 года Управлением исследований и инноваций Сбербанка (модератор А.Р. Ефимов) и МИСиС (см., в частности, панельную дискуссию

Перечислим, следуя И.Н. Вольнову, ряд вызовов, ответить на которые возможно с применением *Art-Science* в образовании. Среди них – снижение возраста потери познавательной активности учащихся и специалистов; фрагментарность картины мира, неопределенность цели, отсутствие оснований; шестой технологический уклад, цифровая экономика, быстрый технологический мир, искусственный интеллект, большие данные, полная автоматизация, цифровой (прозрачный) мир [Вольнов 2019].

В этой связи возникает потребность решения насущных задач. В их числе – обращение к интеллектуальной, и эмоциональной стороне сознания учащихся (эмоциональный интеллект); внесение принципов эстетики (красоты) в проектную деятельность (некрасивые самолеты не летают); красота вносится не на вершине карьерной лестницы (Туполев), а в самом ее начале; искусство как условие возможности уникального, неповторимого, креативного, нестандартного и т.д.; красота как принцип работы с неопределенностью (пространственное мышление). В отношении нового технологического уклада можно утверждска, что искусство становится условием возможности конвергенции (*Art-NBICS*-конвергенция) и трансдисциплинарности (*STEAM- > Art-STEM*).

Полагаем, что итогом такого синтеза станет повышение интереса к образовательному процессу со стороны учащихся, противодействие тренду снижения возраста потери познавательной активности детей; образовательный процесс, построенный не на воспитании усредненных качеств (знаний, навыков, умений), а на выявлении уникальности каждого с возможностью создания индивидуальных образовательных траекторий; развитие не только интеллектуальной компоненты сознания, но и мышления; развитие чувства Красоты как единственного ориентира в мире полной неопределенности и непредсказуемости; адекватный ответ на большие вызовы: ИИ и большие данные, цифровая экономика.

В качестве направлений синтеза науки, искусства и образования могут выступать, например, мобильная робототехника, цифровое производство, нанотехнологии и наноматериалы, полиграфия, для которых программы организованы в виде проектной деятельности. При этом задачи *Art* можно понимать как осмысление цели проекта, работу с общекультурными контекстами, поиск уникальной идеи проекта и его эстетическое воплощение,

---

«STEAM: что добавляет A(rt) в STEM-образование?». – URL https://youtu.be/kothL4ULEp4).

SoftSkills. Задача же *Science/STEM* – техническое воплощение проекта, что предполагает поиск ответа на вопрос о том, как его реализовать.

По мнению В.Н. Вольнова, *Science-Art* воспитывает трансдисциплинарное инженерное диалектическое мышление, способное видеть единство в противоречиях различных контекстов, таких как Наука и Миф, Экономия (наука) – Избыточность (искусство), Массовость (наука) – Уникальность (искусство), Интеллект (ИИ) – Мышление (человек), Сборка мира через цифру (большие данные, ИИ) – Сборка мира через смысл (человек) и т.д. [Вольнов 2018].

STEM характерен для пятого технологического уклада, где основная идея – междисциплинарность. На границе дисциплин возникает новое дисциплинарное знание. Общая картина мира по-прежнему мозаична, т.е. синтез не наблюдается. Происходит введение в массовое образование новых, но усредненных компетенций, работа с интеллектом и информацией. Ключевая идея – польза, эффективность, измеримость, экономия.

Следующая ступень – *STEAM* (*Science-Art*) – имманентна шестому технологическому укладу. Центральная идея – трансдисциплинарность, т.е. выход за дисциплинарные границы. Подразумевается попытка собрать мозаичную картину мира через смыслы. Наблюдается противопоставление смыслов и мышления таким категориям, как информация и интеллект, противопоставление сборке мира ИИ через большие данные – сборке мира человеком через смыслы, отход от усредненных качеств и акцент на уникальности творчества. Очевидной становится попытка найти то, чем человек отличается от ИИ. Главное – креативность, дееспособность в ситуации неопределенности, избыточность мышления. Ввиду трансдисциплинарного подхода *STEAM* может быть преобразован в *STEM-Art* или *Science-Art*. Полюс *ART* выделяется отдельно, как то, что связано со смыслом, мышлением, избыточностью, уникальностью, красотой, и противопоставляется *STEM/Science*: информации, интеллекту, экономии, усреднению и эффективности. В частности, концепция *NBICS*-конвергенции не может быть обеспечена в прагматичном и экономном мышлении. Для ее обеспечения необходим компонент *Art*. В концепции *Art-NBICS* искусство трактуется как трансдисциплинарность и избыточность мышления, что является условием обеспечения технологической конвергенции и инновационного развития.

Заключение

В связи с цифровизацией, развитием искусственного интеллекта и антропоморфной робототехники человек становится творцом новой реальности. Для реализации этого нового статуса требуется развитие человек как целостной творческой личностью. Поэтому на смену процессам нарастания дифференциация научного знания, увеличения разрыва между гуманитарными и естественными науками приходит новая конвергенция. Одна из ее основных составляющих – это сближение науки и искусства как двух основных форм творческой мысли человека. Одним из образцов такого нового синтеза науки и искусства является развивающееся в последнее время направление *STEAM (Science-Art)*. Идея гибридного творчества уже успешно применяется в музыкальных дизайн-проектах, когда композитор или поэт сотрудничают с нейросетью, обучающейся на больших базах музыкальных произведений заданного жанра. Концепция постьюринговой методологии позволяет надеяться, что сложный ИИ, выращенный в гармоничном диалоге с его творцом, будет гарантом сохранения человеческого начала в будущем обществе людей и разумных машин.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Балашова 2018 – *Балашова Е.Ю.* Пространства Бориса Раушенбаха // Психологические исследования. 2018. Т. 11. № 60. С. 2. – URL: http://psystudy.ru/index.php/num/2018v11n60/1596-balashova60.html

Батурин, Крючков 2017 – *Батурин Ю.М., Крючков Б.И.* Двойной портрет: Г.Ю. Оберт и Б.В. Раушенбах // Пилотируемые полеты в космос. 2017. № 2. С. 116–133.

Буданов 1994 – *Буданов В.Г.* Концепция естественнонаучного образования гуманитариев: эволюционно-синергетический подход // Высшее образование в России. 1994. № 4. С. 16–21.

Буданов 2012 – *Буданов В.Г.* Постнеклассические практики и квантовосинергетическая антропология // Постнеклассические практики: опыт концептуализации / под ред. В.И. Аршинова, О.Н. Астафьевой. – СПб.: МИРЪ, 2012. С. 37–62.

Буданов 2017 – *Буданов В.Г.* Методология синергетики в постнеклассической науке и в образовании / 4-е изд., доп. – М.: УРСС, 2017.

Буданов, Синицына 2020 – *Буданов В.Г., Синицына Т.А.* Квантово-синергетическая онтология обобщенной телесности: От антропологии театра к очеловечиванию искусственного интеллекта, проблема границ // Культура и искусство. 2020. № 7. С. 13–28.

Вольнов 2018 – *Вольнов И.Н.* Единство науки и искусства как ответ на вызовы цифровой реальности // Контуры цифровой реальности: Гуманитарно-технологическая революция и выбор будущего / под.

ред. В.В. Иванова, Г.Г. Малинецкого, С.Н. Сиренко. – М.: Ленанд, 2018. С. 320–326.

Вольнов 2019 – *Вольнов И.Н.* Смыслы в цифровую эпоху. Человек сингулярный // Обсерватория культуры. 2019. Т. 16. № 4. С. 340–348.

Сноу 1973 – *Сноу Ч.П.* Две культуры и научная революция // *Сноу Ч.П.* Две культуры: сборник публицистических работ. – М.: Прогресс, 1973. С. 18–61.

Степин 2000 – *Степин В.С.* Теоретическое знание. – М.: Прогресс-Традиция, 2000.

Степин, Кузнецова 1994 – *Степин В.С., Кузнецова Л.Ф.* Научная картина мира в культуре техногенной цивилизации. – М.: ИФ РАН, 1994.

Щедровицкий 1997 – *Щедровицкий Г. П.* Философия. Наука. Методология. – М.: Школа культурной политики, 1997.

Ballora 2014 – *Ballora M.* Sonification, Science and Popular Music: In Search of the 'Wow' // Organised Sound. 2014. Vol. 19. No. 1. P. 30–40.

Chursinova, Stebelska 2021 – *Chursinova O., Stebelska O.* Is the Realization of the Emotional Artificial Intelligence Possible? Philosophical and Methodological Analysis // Filosofija. Sociologija. 2021. Vol. 32. No. 1. P. 76–83.

Efimov 2020 – *Efimov A.* Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence // Artificial General Intelligence. 13th International Conference, AGI 2020 (Lecture Notes in Computer Science. Vol. 12177) / ed. by B. Goertzel, A. Panov, A. Potapov, R. Yampolskiy. – Cham: Springer, 2020. P. 83–94.

Ginter, Heuer, Franzobel 2011 – *Ginter P., Heuer R.-D., Franzobel.* LHC – Large Hadron Collider. – Baden: Edition Lammerhuber, 2011.

Quinton, McGregor, Benyon 2016 – *Quinton M., McGregor I., Benyon D.* Sonifying the Solar System // Sonic Information Design: Proceedings of the 22nd Annual International Conference on Auditory Display / ed. by D. Worrall. – Canberra: School of Music, Australian National University, 2016. P. 28–35.

Schuller, Schuller 2018 – *Schuller D., Schuller B.W.* The Age of Artificial Emotional Intelligence // Computer. 2018. Vol. 51. No. 9. P. 38–46.

Schwärzler 2016 – *Schwärzler M.* At Face Value and Beyond: Photographic Constructions of Reality. – Bielefeld: Transcript-Verlag, 2016.

Yonck 2020 – *Yonck R.* Heart of the Machine: Our Future in a World of Artificial Emotional Intelligence. – New York: Arcade Publishing, 2020.

REFERENCES

Balashova E.Yu. (2018) Spaces of Boris Rauschenbach. *Psikhologicheskie issledovaniya – Psychological Studies.* Vol. 11, no. 60, p. 2. Retrieved from http://psystudy.ru/index.php/num/2018v11n60/1596-balashova60.html

Ballora M. (2014) Sonification, Science and Popular Music: In Search of the 'Wow'. *Organised Sound.* Vol. 19, no. 1, pp. 30–40.

Baturin Yu.M. & Kryuchkov B.I. (2017) Double Portrait: G.Yu. Obert and B.V. Rauschenbach. *Pilotiruemye polety v kosmos.* No. 2, pp. 116–133.

Budanov V.G. (1994) The Conception of Natural Science Education for Humanities: An Evolutionary-Synergetic Approach. *Vysshee obrazovanie v Rossii.* No. 4, pp. 16–21 (in Russian).

Budanov V.G. (2012) Post-Non-Classical Practices and Quantum-Synergetic Anthropology. In: Arshinov V.I. & Astafyeva O.N. (Eds.) *Post-Non-Classical Practices: The Experience of Conceptualization* (pp. 37–62). Saint Petersburg: Mir (in Russian).

Budanov V.G. (2017) *Methodology of Synergetics in the Non-Classical Science and in Education* (4th ed.). Moscow: URSS (in Russian).

Budanov V.G. & Sinitsyna T.A. (2020) Quantum-Synergetic Ontology of Generalized Corporeality: From Anthropology of Theater to Humanization of Artificial Intelligence, the Problem of Boundaries. *Kul'tura i iskusstvo.* No. 7, pp. 13–28. (in Russian).

Chursinova O. & Stebelska O. (2021) Is the Realization of the Emotional Artificial Intelligence Possible? Philosophical and Methodological Analysis. *Filosofija. Sociologija.* Vol. 32, no. 1, pp. 76–83.

Efimov A. (2020) Post-Turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence. In: Goertzel B., Panov A., Potapov A., & Yampolskiy R. (Eds.) *Artificial General Intelligence. 13th International Conference, AGI 2020* (Lecture Notes in Computer Science. Vol. 12177) (pp. 83–94). Cham: Springer.

Ginter P., Heuer R.-D., & Franzobel (2011) *LHC – Large Hadron Collider.* Baden: Edition Lammerhuber.

Quinton M., McGregor I., & Benyon D. (2016) Sonifying the Solar System. In: Worrall D. (Ed.) *Sonic Information Design: Proceedings of the 22nd Annual International Conference on Auditory Display* (pp. 28–35). Canberra: School of Music, Australian National University.

Schuller D. & Schuller B.W. (2018) The Age of Artificial Emotional Intelligence. *Computer.* Vol. 51, no. 9, pp. 38–46.

Schwärzler M. (2016) *At Face Value and Beyond: Photographic Constructions of Reality.* Bielefeld: Transcript-Verlag.

Shchedrovitsky G.P. (1996) *Philosophy. Science. Methodology.* Moscow: Shkola kul'turnoy politiki (in Russian).

Snow C.P. (1959) *The Two Cultures and the Scientific Revolution.* Cambridge, UK: Cambridge University Press (Russian translation: Moscow: Progress, 1977).

Stepin V.S. (2000) *Theoretical Knowledge.* Moscow: Progress-Traditsiya (in Russian).

Stepin V.S. & Kuznetsova L.F. (1994) *Scientific Picture of the World in the Culture of Technogenic Civilization*. Moscow: RAS Institute of Philosophy (in Russian).

Volnov I.N. (2018) The Unity of Science and Art as a Response to the Challenges of Digital Reality. In: Ivanov V.V., Malineckiy G.G., & Sirenko S.N. (Eds.) *The Contours of Digital Reality: Humanitarian and Technological Revolution and the Choice of the Future* (pp. 320–326). Moscow: Lenand (in Russian).

Volnov I.N. (2019) Meanings in the Digital Age. Singular Man. *Observatoriya kultury.* Vol. 16, no. 4, pp. 340–348. (in Russian).

Yonck R. (2020) *Heart of the Machine: Our Future in a World of Artificial Emotional Intelligence.* New York: Arcade Publishing.

DOI: 10.30727/0235-1188-2021-64-1-134-148
Оригинальная исследовательская статья
Original research paper

Искусственный интеллект: как углеродная жизнь создала кремниевую

А.Ю. Сторожук
Институт философии и права Сибирского отделения РАН, Новосибирск, Россия

Аннотация

В статье рассматриваются основные этапы развития искусственного интеллекта (ИИ). Цель статьи – выявить влияние компьютеризации на общественную жизнь и мировоззрение в настоящее время и в будущем. Метод исторического анализа позволил выделить этапы развития нейросетей, провести сравнение прогнозов и реальных достижений в области моделирования ИИ. Охарактеризованы три главных периода: «первая весна», «вторая весна» и «зима» эры ИИ (период стагнации). Во время первых обнадеживающих разработок возникли трудности технического характера. На первом этапе разработчики пытались копировать природу, т.е. нейроны естественного мозга, что оказалось слишком сложным. Кроме того, первые алгоритмы обучения нейронных сетей сопровождались принципиальными трудностями, такими как экспоненциальное затухание вектора ошибки при распространении во внутренние слои нейронной сети. Потребовались годы для нахождения новых алгоритмических решений, накопления данных для обучения нейронных сетей, совершения следующего технологического прорыва, известного как «вторая весна» ИИ. В процессе и по итогам анализа установлено, что прогнозы возможностей ИИ часто были более скромными, чем реальные достижения. В заключение в статье сделан новый прогноз, согласно которому вычислительная техника будет играть не только инструментальную роль, но и окажет существенное влияние на мировоззренческие ценности общества. Это произойдет как по причине усиления роботизации, так и за счет вытеснения машинами людей творческих профессий, что усилит сегрегацию общества и тенденции к индивидуализму. Проблема реализма и поиск критериев истины будет также обостряться по мере развития самопрограммируемых нейросетей.

Ключевые слова: философия искусственного интеллекта, нейронные сети, информация, технический прогресс, мышление, социальная философия, общество.

Анна Юрьевна Сторожук – доктор философских наук, ведущий научный сотрудник Института философии и права Сибирского отделения РАН.

stor71@mail.ru

http://orcid.org/0000-0002-0893-6212

Для цитирования: *Сторожук А.Ю.* Искусственный интеллект: как углеродная жизнь создала кремниевую // Философские науки. 2021. Т. 64. № 1. С. 134–148. DOI: 10.30727/0235-1188-2021-64-1-134-148

Artificial Intelligence: How Carbon-Based Life Has Created Silicon-Based Life

A.Yu. Storozhuk
Institute of Philosophy and Law, Siberian Branch,
Russian Academy of Sciences, Novosibirsk, Russia

Abstract

The paper considers the main stages of the development of artificial intelligence (AI). The aim of the paper is to reveal the influence of computerization on social life at present and in the future. The historical analysis method was used to identify the main stages in the development of neural networks and to compare some predictions and real achievements in the field of AI modeling. The article identifies three main periods, widely known as the first and second spring of the era of artificial intelligence and the "winter of AI" – a period of stagnation. The earliest promising developments ran into technical difficulties. At the first stage, the developers tried to copy nature, that is, the neurons of the natural brain, and the tasks turned out to be too complicated. In addition, the first algorithms for training neural networks encountered fundamental difficulties, such as exponential decay of the error vector when propagating into the inner layers of the neural network. It took years to find new algorithmic solutions, to accumulate data for training neural networks, for the next technological breakthrough, known as the "second spring of AI." The result of the analysis showed that predictions of AI capabilities were often more modest than actual achievements. In the present paper, we make new predictions, according to which computer technology is going play not only an instrumental role but will also have a significant impact on the ideological values of society. This will happen not only due to increasing automation but also due to substitution of creative professions by machines, which will increase the segregation of society and the trend towards individualism. The problem of realism and the search for criteria of truth will also gain importance as self-programmable neural networks evolve.

Keywords: philosophy of artificial intelligence, neural networks, information, technical progress, thinking, social philosophy, society.

Anna Yu. Storozhuk – D.Sc. in Philosophy, Leading Research Fellow, Institute of Philosophy and Law, Siberian Branch, Russian Academy of Sciences.
stor71@mail.ru
http://orcid.org/0000-0002-0893-6212

For citation: Storozhuk A.Yu. (2021) Artificial Intelligence: How Carbon-Based Life Has Created Silicon-Based Life. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 134–148.
DOI: 10.30727/0235-1188-2021-64-1-134-148

Введение

Термин «искусственный интеллект» появился в середине XX века, когда обсуждалась теоретическая возможность построить мыслящие машины. Предполагалось, что они смогут заменить человека во вредных условиях (шахта, космос) или будут выполнять монотонную работу, например, отбраковку деталей на заводах. При этом машине отводилась подчиненная роль послушного помощника.

Положительные ответы на вопрос «Может ли машина мыслить?» разделили на два класса. Речь идет о сильном и слабом искусственном интеллекте (ИИ). Последний был самым распространенным, т.к. предполагал лишь частичное превосходство машины в некоторых областях, таких как счет, память [Безлепкин 2020], поиск данных и т.п., что стало очевидным уже на ранних этапах развития вычислительной техники. Сильный ИИ имел немного сторонников, поскольку вплоть до начала нынешнего века не допускали мысли о возможном превосходстве машины над человеком в творчестве, интуиции, умении открывать новое знание [Чалмерс 2013]. Полагали, что машина не производит нового знания, но лишь быстро ищет в базе данных имеющиеся варианты или быстро считает [Фарберович 2019]. Кроме того, считалось невозможным для ИИ решение задач, требующих бесконечного перебора, в условиях действия непредсказуемых факторов и т.п. Например, нельзя не вспомнить о том, как компьютерная программа обыграла в шахматы чемпиона мира (шахматный суперкомпьютер *Deep Blue* в 1997 году выиграл у чемпиона мира Гарри Каспарова). Хотя знатоки утверждали, что компьютер никогда не

выиграет у человека, т.к. данная игра требует бесконечного перебора вариантов. По состоянию на 2016 год программа *AlphaGo Zero* с большим отрывом обыграла всех чемпионов мира и национальных чемпионов [Silver, Huang, Maddison et al. 2016].

Насколько оправдались ожидания и долговременные прогнозы, сделанные десятилетия назад? Каковы сегодня отношения человека и машины? Чего можно ожидать в ближайшее время? Каково влияние компьютеризации на общество? Как компьютеры изменили отношения между людьми?

Времена года: краткая история искусственного интеллекта

Сегодняшние дни принято называть «второй весной» ИИ. «Первой весной» именуют период энтузиазма, охвативший научное сообщество в середине XX века, когда появились первые алгоритмы и механические устройства, подобные «черепашке» Грея Уолтера [Walter 1950]. Волна энтузиазма нашла на камень трудностей технического характера, которые были непреодолимы во второй половине прошлого века, известного как «зима» ИИ.

Первые попытки создания искусственного разума представляли собой копирование природы. Архитектура искусственного нейрона создавалась по образу и подобию естественной нервной клетки, первая компьютерная нейронная сеть пыталась подражать нейронным сетям животных [Коннова 2020]. Подражание служило бледным подобием, т.к. естественная нервная клетка может иметь до двадцати тысяч синапсов, соединяющих ее с другими клетками. Кроме того, известно о перцептивных полях, т.е. объединениях нервных клеток сетчатки в определенные структуры. По образу и подобию организации нервных клеток в сетчатке кошки созданы первые искусственные нейронные сети, самой известной из которых является перцептон Розенблатта [Rosenblatt 1961]. Он представлял собой массив, состоящий из входного слоя, одного внутреннего слоя (т.е. однослойной нейронной сети) и выходного нейрона, куда записывались результаты вычислений. Каждый нейрон входного слоя был связан со всеми нейронами внутреннего слоя, который и производил вычисления. Часто это выглядело как суммирование значений входных нейронов, умноженных на коэффициент, который определял т.н. вес данной связи. Перцептон был обучаемой нейронной сетью. Обучение относилось к корректировке весов, т.е. на изменения коэффициентов умножали входные данные. Чтобы произвести корректировку весов, резуль-

тат нейронной сети сравнивали с заранее известным правильным ответом и вычисляли вектор ошибок, в зависимости от величины которого производили корректировку весов.

Очень быстро начались проблемы технического характера. Так, для однослойного перцептона не была доказана теорема сходимости в отношении некоторых логических операций [Minsky, Papert 1969]. Кроме того, его мощности не хватало для решения сложных задач. Появились сети с большим числом внутренних слоев. Но у многослойного перцептона возникли проблемы с обучением, т.к. вектор ошибки при распространении внутрь затухал экспоненциально, т.е. обучались практически один-два выходных слоя.

Как могла наступить «вторая весна»? Выделим три источника или составных части: появление новых алгоритмов, накопление больших массивов данных, возникновение новых вычислительных мощностей. Новые алгоритмы были нацелены на улучшение обучаемости нейронных сетей. Последние имели преимущества перед жестко заданными алгоритмами в своей гибкости. Нейронные сети позволяли параллельно обрабатывать массивы информации. В частности, наиболее известны рекурсивные нейронные сети, которые способствовали распространению вектора ошибок на внутренние слои [Rumelhart, Hinton, Williams 1986]. Нейронные сети обладали высокой помехоустойчивостью: нарушения приводили к снижению их производительности, но не к полному отказу.

Об этом писал и Н. Бостром: «Важно, что нейронные сети представляли собой самообучающиеся интеллектуальные системы, т.е. накапливали опыт, умели делать выводы из обобщенных примеров и находить скрытые статистические образы во вводимых данных. Это делало сети хорошим инструментом для решения задач классификации и распознавания образов. Например, создав определенный набор сигнальных данных, можно было обучить нейронную сеть воспринимать и распознавать акустические особенности подводных лодок, мин и морских обитателей с большей точностью» [Бостром 2016, 20]. Широкое распространение получил блочный подход – конструктор нейронных сетей, т.е. ряд блоков, из которых можно собирать нейронные сети [Варламов 2011]. Развитие последних стало возможным благодаря новым методам представления знаний в компьютерах как ядра ИИ. Методы поиска решений носят в основном переборный характер. Но в различных представлениях одна и та же задача может иметь разный

размер. Более компактное представление позволяет решать более сложные задачи при аналогичных вычислительных мощностях [Новиков 2019, 11].

Существенным условием, обеспечившим очередной рывок в технологиях ИИ, т.е. наступление «второй весны» ИИ, стало появление новых компьютерных технологий – методов эволюционных вычислений, таких как генетический алгоритм. Бостром объяснял происходящее следующим образом: «В эволюционных моделях в первую очередь создаются начальные популяции тех или иных решений (могут быть либо структуры данных, либо программы обработки данных), затем – в результате случайной мутации и размножения (“скрещивания”) имеющихся популяций – генерируются новые популяции. Периодически вследствие применения критерия отбора (по наличию целевой функции, или функции пригодности) количество популяций сокращается, что позволяет войти в новое поколение лишь лучшим решениям-кандидатам. В ходе тысяч итераций среднее качество решений в популяции постепенно повышается» [Бостром 2016, 20]. Генетические алгоритмы используют идею естественного отбора Дарвина и позволяют получать самые «жизнеспособные» продуктивные программы.

Огромную роль в приближении «второй весны» сыграло накопление большого количества данных. Для обучения нейронной сети требуется много образцов[1]. На начальном этапе развития ИИ трудно было обеспечить такое количество. Сегодня, благодаря сети Интернет и социальным сетям, у многих пользователей появилась возможность выкладывать собственные изображения и тексты. Это обеспечило появление объема данных, достаточного для качественного обучения нейронных сетей.

Важную роль сыграло и увеличение вычислительных мощностей. Благодаря видеоиграм и криптовалютам получили широкое распространение скоростные процессоры с мощными видеокартами. Некоторые операции, например, сортировка, быстрее происходят на видеокартах, чем на процессорах. Это оказалось решающим фактором, т.к. обучение и работа нейросетей требует больших вычислительных мощностей. В итоге к 2020 году, по оценке М.С. Бурцева, если сравнивать с живой природой по коли-

[1] Например, чтобы научить нейронную сеть отличать кошку от собаки, необходимо около миллиона изображений. (См.: *Хуршудов А.* Про котиков, собак, машинное обучение и deep learning // Хабр. 2014. 16 июня. – URL: https://habr.com/ru/post/226347/)

честву нейронов, то общий уровень ИИ сравним с интеллектом мыши [Соколенко 2020].

Влияние прогресса технологий ИИ на развитие общества

Еще несколько десятилетий назад Р. Пенроуз в работе «Новый ум короля» (1989) писал, что компьютеры никогда не смогут заменить человека, поскольку у них нет понимания того, что они делают [Пенроуз 2020, 33]. По аналогии с мысленным экспериментом «китайская комната» Серла [Searle 2001] Пенроуз заключает, что нейросеть может быть хорошей имитацией, но она никогда не достигнет понимания. Кроме того, некоторые невербальные способы мышления не алгоритмизуемы [Пенроуз 2020, 338–344].

Однако сегодня машины прошли тест Тьюринга (случай испытания чат-ботов описан в 2014 году [Новиков 2019, 21]). Давно создана нейросеть, которая работает как психотерапевт [Weizenbaum 1966]. Нейросети занимаются творчеством. Они пишут статьи, сочиняют музыку, рисуют, декламируют и даже программируют другие нейросети, т.е. вытесняют человека из ряда типичных творческих специальностей.

В ранний период появления ИИ считали, что машина лишь комбинирует имеющиеся знания и не способна породить ничего нового. Однако, когда нейросеть выиграла у чемпиона мира партию в го [Metz 2016], специалисты, анализирующие партию, обнаружили непонятные ходы, сделанные нейросетью. Смысл этих ходов стал ясен только к середине партии. Это была позиционная игра, которая и позволила нейросети выиграть. Интересен тот факт, что подобные приемы в го ранее не использовались и не были известны. Профессиональные игроки в го нашли ходы нейросети очень изящными. Если говорить о стратегии именно этой игры, то машина в данном случае создала новое.

Еще одна задача для нейросети – научиться рисовать «под Ван Гога» [Gatys, Ecker, Bethge 2020]. В целях ее решения использовали две нейросети. Задача для одной из них состояла в том, чтобы на основе анализа картин Ван Гога научиться переделывать фотографии в стиле его живописи. Другая нейросеть должна была установить, подлинная ли картина. В результате совместного обучения первая нейросеть выдавала такие результаты, что человек уже не мог отличить подлинник от подделки. Иными словами, рисунок нейросети предстал таким, который мог бы принадлежать Ван Гогу. Только знание того, какие картины

были им написаны при жизни, позволило экспертам отличить подлинник от т.н. компьютерной подделки.

В настоящее время, на наш взгляд, влияние компьютеров на общество приобрело экзистенциальный характер. Компьютер дает возможность суррогатной жизни подобно тому, как соевые продукты представляют собой суррогатное мясо. Прежде всего важной частью является суррогатное общение. Усиливается индивидуализация общества, подогреваемая, помимо остального, активно насаждаемой моралью потребительского общества [Santos, Varnum, Grossmann 2017].

Существенное воздействие на психику и социальное поведение подростков оказывают также компьютерные игры, в которых «герой обычно не учитывает чувства или интересы других персонажей и демонстрирует утилитарное отношение к другим» [Noga 2016, 578]. Наблюдается разница в социальном поведении и системе ценностей между подростками, увлеченными играми с элементами агрессивности и жестокости, и подростками, которые в них не играют. Влияние медиакультуры на общество исследуется в специальных работах. Обратим внимание лишь на философски значимые аспекты, такие как влияние на формирование мировоззрения и новых стилей когнитивной активности, рост социального контроля, изменение системы ценностей и влияние на восприятие мира, понимание реальности [Perse, Lambe 2016, 1].

С развитием компьютерных технологий повышается точность воспроизведения образов. Ряд изображений, созданных нейросетями, люди уже не могут отличить от настоящих. Произошли серьезные изменения в кинематографе: во многих жанрах уже практически не снимаются картины без широкого использования компьютерной графики. На первый взгляд представляется, что компьютерное моделирование способствует повышению чувства реализма. Если проводить аналогию с развитием фотографии, то нельзя не отметить, что сначала, при переходе от черно-белых фотографий к цветным, степень реализма повышалась. Фотоискусство носило документальный характер, играло роль свидетельства, подтверждающего истинность словесной истории. С развитием программ для обработки изображений, таких как *Photoshop*, фотография теряет роль объективного свидетельства, поскольку подделка становится порой очень качественной, настолько, что ее сложно отличить от реально сделанной фотографии. Сегодня появляются нейросети, способные обрабатывать

фотографии и путем их комбинации получать практически любые изображения.

Тема реализма тесно связана с понятием истинности. В контексте научных теорий различали две основные теории: когерентную, предусматривающую отсутствие внутреннего противоречия, и корреспондентную, базирующуюся на идее о соответствии внешнему миру. В настоящее время, когда статья может быть написана программой, генерирующей тексты, а доступ к документальным свидетельствам опосредован компьютером, проблема отличия истины от лжи может оказаться неразрешимой.

Изначально техника создавалась, чтобы расширить возможности людей и облегчить им труд. Но в итоге такое расширение возможностей с помощью компьютерных технологий привело к подчиненному положению людей. Происходит замещение человека автоматикой. Люди, которые остаются, играют роль обслуживающего персонала, т.е. они нужны, чтобы помогать машинам, а не для того, чтобы машина помогала им [Маркофф 2016, 78]. Если это происходит сегодня, то ожидаемое потенциальное взрывоподобное развитие интеллекта, когда алгоритмы самостоятельно будут писать новое, более совершенное, данный процесс только усилит.

Заключение

В основе проектирования нейронных сетей, иных видов алгоритмов, актуальных при создании ИИ, находится подражание устройству живых организмов. Однако это подражание, по сути, происходит лишь в одном аспекте – информационном. По аналогии с естественной памятью создаются условия для хранения информации, по аналогии с естественной коммуникацией происходит передача, обработка, защита от помех информации в вычислительных машинах. Скорость технического прогресса намного опережает скорость биологической эволюции. Информационные технологии прогрессируют стремительно, намного медленнее развивается мир животных и растений. Многие обеспокоены тем, что человечество утратит пальму первенства в вопросе доминирования над другими биологическими видами, а также займет подчиненную роль, сравнимую с местом домашних животных [Butler 1863]. Финальной аналогией мы обязаны книге Ч. Докинза «Эгоистичный ген» [Докинз 2013]. Если мы заменим слово «ген» словом «информация», то, в сущности, создадим усло-

вия для более быстрого и широкого распространения последней. Если биологические существа, состоящие из белков, копировали и распространяли информацию в ограниченных масштабах, то полупроводниковые устройства намного превосходят углеводородные по скорости копирования и эволюции.

Фактически мы сами создаем преемников. Речь идет о следующей цивилизации, но уже на кремниевой основе. Основные типы прогнозов относительно будущего человечества в атмосфере доминирования ИИ можно разделить условно на три класса.

1. Оптимистичные сценарии, предполагающие улучшение условий жизни человека. Первая функция ИИ, как и любых других вычислительных мощностей, состоит в том, чтобы быть локомотивом экономики. Анализ цивилизаций за последние две тысячи лет показал [Колташов 2018], что причины экономических кризисов всегда были одними и теми же – падение покупательского спроса и истощение сельскохозяйственных земель. Выход из кризиса тоже был принципиально одинаковым, он связан с техническими новшествами, разогревающими спрос на рынках. Компьютерная индустрия является особенно значимым сегментом мировой экономики и практически единственным, показавшим рост в условиях эпидемии COVID-19. Взрыв спроса произошел на видеоигры, программы для дистанционного общения, проведения конференций и др. Тем более что данный сегмент рынка не является ресурсоемким: в основе вычислительной техники лежат полупроводниковые микросхемы, а кремний – третий по распространенности элемент земной коры, т.е. сырье для изготовления полупроводниковой техники общедоступно. Информация служит, возможно, единственным ресурсом многоразового использования, остальные ресурсы расходуются при пользовании ими, сокращаются. Т.н. развитые страны занимаются производством, скорее, моделей производства, которые воплощаются в других странах с индустриальными обществами [Федотова 2016, 185].

Ожидается значительное улучшение жизни людей как за счет оптимизации социальной организации, так и за счет свободного доступа к информационным ресурсам. Дж. Нэсбит и П. Абурден в исследовании «Мегатенденции: десять новых направлений, преобразующих нашу жизнь» предсказали, что произойдет смена модели управления обществом: переход от централизованной иерархической структуры к децентрализованной сетевой структуре [Naisbitt, Aburdene 1991].

2. Негативные сценарии, связанные с подчинением людей машинам. В настоящее время ИИ во многих сферах интеллектуальной деятельности превосходит возможности человека. Неслучайно возникает следующий вопрос. Сложный для понимания ИИ уже сегодня, станет ли он понятнее через пять или сто лет? Предполагается, что одновременно со сложностью и превосходством будет расти и могущество машин, а значит, и способность влиять на жизнь людей. Негативные сценарии основаны на приписывании машинам человеческих слабостей: жажды денег и власти, тщеславия, иных. Насколько реалистичны такие тенденции?

Многие авторы, начиная с фантаста А. Азимова, писавшего о трех законах роботехники, указывают на необходимость передачи машинам моральных норм. В этом видится антропоморфизация техники: именно моральные нормы и совесть являются сдерживающей силой естественного интеллекта. Но как «воспитать» машину? Н. Бостром делает акцент на трудности формализации моральных правил, а также на том, что «зрелый», т.е. прошедший обучение ИИ, не склонен изменяться. Следовательно, моральные правила и систему запретов нужно закладывать изначально *in silico* [*Бостром* 2016].

Представляется надуманным приписывание технике человеческой мотивации. Тем не менее даже при ее отсутствии техника способна негативно влиять на жизнь людей. Вероятно, будет продолжаться вытеснение людей из ряда интеллектуальных и творческих профессий. Ранее существовало мнение о том, что машины заменят человека на тяжелых работах во вредных условиях. Однако стоимость оборудования для работы нейросети такова, что ее не отнести к вредным условиям. Скорее, человек самостоятельно пойдет в шахту добывать уголь, чтобы удовлетворить растущие потребности компьютеров в электроэнергии. При появлении первых вычислительных машин возник целый штат сотрудников по их обслуживанию: программисты, операторы, электрики, служба профилактики и технического обслуживания. Эта тенденция станет еще более активна. В сферу обслуживания машин будет вовлечено все больше людей.

3. Радикальное изменение человеческого и социального. Видение в черно-белых тонах является слишком примитивным пониманием истории. На наш взгляд, ни одну тенденцию нельзя охарактеризовать однозначно в позитивном или негативном ключе.

Но техника окажет и уже оказывает значительное влияние на жизнь человека, в чем-то облегчая его существование, а в чем-то требуя уступок и жертв.

Тенденции развития общества могут оказаться разнонаправленными: с одной стороны, происходит объединение человечества в информационном пространстве с медиакоммуникациями, позволяющими осуществлять общий контроль, вести пропаганду универсальных ценностей (по принципу равенства всех перед всеми). С другой стороны, как нам кажется, равенства не будет, развитие ИИ приведет к значительному расслоению общества и его дифференциации. Менее образованные будут занимать место пользователей и обслуживающего персонала, наиболее способные – выполнять функции программистов и наладчиков оборудования. Учитывая экспоненциально возрастающую скорость технического прогресса и тем, и другим придется постоянно учиться, чтобы сохранить свой *status quo*. Постоянное обучение в аспекте технической составляющей окажет влияние и на структуру сознания. Структурные и математические способности будут развиваться, творческие «правополушарные» способности и интуиция окажутся «за бортом» мейнстрима. Цифровая жизнь подчинит себе жизнь экономическую и политическую, а повсеместность цифровых технологий будет обеспечивать универсальность и унифицированность мышления, системы ценностей, что создаст условия для концентрации контроля и власти в одних руках.

Свобода будет широко пропагандироваться, но в действительности свобода перемещения, выбора профессии, самореализации, поиска себя будут в значительной степени ограничены. Произойдет существенное ослабление авторитета мудрых людей, т.к. техническая грамотность заменит мудрость, а жизненный опыт окажется не применимым в быстро изменяющихся условиях и технологиях. Кроме того, виртуальное (суррогатное) общение будет способствовать потере навыков реального общения, а значит, и развитию индивидуализма. Даже находясь в обществе, человек будущего не сможет продуктивно общаться, не станет членом коллектива. Рано или поздно эти эффекты, вызванные развитием техники, приведут к отказу от традиционных общественных ценностей (любви, дружбы, взаимовыручки), от чувства коллективизма. В этом проявляются, полагаем, основные риски цифровизации.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Безлепкин 2020 – *Безлепкин Е.А.* Сообщество разума: идеи М. Минского об интеллектуальных агентах // Философия науки. 2020. № 3 (86). С. 105–118.

Бостром 2016 – *Бостром Н.* Искусственный интеллект. Этапы. Угрозы. Стратегии. – М.: Манн, Иванов и Фербер, 2016.

Варламов 2011 – *Варламов О.* Миварный подход к разработке интеллектуальных систем и проект создания мультипредметной активной миварной интернет-энциклопедии // Известия Кабардино-Балкарского научного центра РАН. 2011. № 1 (39). С. 55–64.

Докинз 2013 – *Докинз Р.* Эгоистичный ген. – М.: АСТ, CORPUS, 2013.

Колташов 2018 – *Колташов В.* Экономические кризисы с III века до наших дней: цикличность, природа и логика современных потрясений // Логос. 2018. Т. 28. № 6 (127). С. 27–72.

Коннова 2020 – *Коннова Н. С.* Нейронные сети, или Как обучить искусственный интеллект // Интернет изнутри. 2020. № 3. С. 38–51.

Маркофф 2016 – *Маркофф Д.* Homo Roboticus? Люди и машины в поисках взаимопонимания. – М.: Альпина нон-фикшн, 2016.

Новиков 2019 – *Новиков Ф.А.* Символический искусственный интеллект: математические основы представления знаний. – М.: Юрайт, 2019.

Пенроуз 2020 – *Пенроуз Р.* Новый ум короля: о компьютерах, мышлении и законах физики. – М.: URSS, Ленанд, 2020.

Соколенко 2020 – *Соколенко О.* «Сегодня общий уровень ИИ сопоставим с уровнем интеллекта мыши или пчелы. Боимся ли мы, что мыши и пчелы захотят захватить мир?» Заведующий лабораторией нейронных систем и глубокого обучения МФТИ Михаил Бурцев – о рисках искусственного интеллекта для человечества // Republic. 2020. 25 сентября. – URL: https://republic.ru/posts/97895

Федотова 2016 – *Федотова В.Г.* Модернизация и культура. – М.: Прогресс-Традиция, 2016.

Чалмерс 2013 – *Чалмерс Д.* Сознающий ум. В поисках фундаментальной теории. – М.: Либроком, 2013.

Butler 1863 – *Butler S.* Darwin Among the Machines // The Press. 1863. June 13. P. 180–185.

Gatys, Ecker, Bethge 2020 – *Gatys L., Ecker A., Bethge M.* A Neural Algorithm of Artistic Style // Journal of Vision. 2016. Vol. 16. No. 12. P. 1–16.

Metz 2016 – *Metz C.* Why the Final Game Between AlphaGo and Lee Sedol Is Such a Big Deal for Humanity // Wired News. 2016. March 14. – URL: https://www.wired.com/2016/03/final-game-alphago-lee-sedol-big-deal-humanity/

Minsky, Papert 1969 – *Minsky M., Papert S.* Perceptrons. – Cambridge, MA: MIT Press, 1969.

Naisbitt, Aburdene 1990 – *Naisbitt J., Aburdene P.* Megatrends 2000. Ten New Directions for the 1990's. – New York: William Morrow, 1990.

Noga 2016 – *Noga H.* Selected Factors of the Socialization Impact of Computer Games and Their Content // Society. Integration. Education: Proceedings of the International Scientific Conference. 2016. Vol. 2. P. 569–578.

Perse, Lambe 2016 – *Perse E.M., Lambe J.* Media Effects and Society. – New York: Routledge, 2016.

Rosenblatt 1961 – *Rosenblatt F.* Principles of Neurodynamics. – Washington, DC: Spartan Books, 1961.

Rumelhart, Hinton, Williams 1986 – *Rumelhart D.E., Hinton G.E., Williams R.J.* Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations. – Cambridge, MA: MIT Press, 1986.

Santos, Varnum, Grossmann 2017 – *Santos H.C., Varnum M.E.W., Grossmann I.* Global Increases in Individualism // Psychological Science. 2017. Vol. 28. No. 9. P. 1228–1239.

Searle 2001 – *Searle J.* Chinese Room Argument // The MIT Encyclopedia of the Cognitive Sciences / ed. by R.A. Wilson, F.C. Keil. – Cambridge, MA: MIT Press, 2001. P. 115–116.

Silver, Huang, Maddison et al. 2016 – *Silver D., Huang A., Maddison C. et al*. Mastering the Game of Go with Deep Neural Networks and Tree Search // Nature. 2016. Vol. 529. No. 7587. P. 484–489.

Walter 1950 – *Walter W.G*. An Imitation of Life // Scientific American. 1950. Vol. 182. No. 5. P. 42–45.

Weizenbaum 1966 – *Weizenbaum J.* ELIZA – a Computer Program for the Study of Natural Language Communication between Man and Machine // Communications of the ACM. 1966. Vol. 9. No. 1. P. 36–45.

REFERENCES

Bezlepkin E. A. (2020) The Community of Reason: M. Minsky's Ideas about Intellectual Agents. *Filosofiya nauki.* No. 3, pp. 105–118 (in Russian).

Bostrom N. (2014) *Superintelligence: Paths, Dangers, Strategies.* Oxford: Oxford University Press (Russian translation: Moscow: Mann, Ivanov and Ferber, 2016).

Butler S. (1863, June 13) Darwin Among the Machines. *The Press.* Pp. 180–185.

Chalmers D. (1996) *The Conscious Mind. In Search of a Fundamental Theory.* New York: Oxford University Press (Russian translation: Moscow: Linrokom, 2013).

Dawkins R. (1976) *The Selfish Gene.* Oxford: Oxford University Press (Russian translation: Moscow: AST, Corpus, 2013).

Fedotova V.G. (2016) *Modernization and Culture.* Moscow: Progress-Traditsiya (in Russian).

Gatys L., Ecker A., & Bethge M. (2020) A Neural Algorithm of Artistic Style. *Journal of Vision.* Vol. 16, no. 12, pp. 1–16.

Koltashov V. (2018) Economic Crises from the Third Century to the Present Day: Cyclicality, Nature and Logic of Modern Upheavals. *Logos.* Vol. 28, no. 6, pp. 27–72 (in Russian).

Konnova N.S. (2020) Neural Networks, or How to Train Artificial Intelligence. *Internet iznutri.* No. 3, pp. 38–51 (in Russian).

Markoff D. (2015) *Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots.* New York: Ecco Books (Russian translation: Alpina non-fiction, 2016).

Metz C. (2016, March 14) Why the Final Game Between AlphaGo and Lee Sedol Is Such a Big Deal for Humanity. *Wired News.* Retrieved from https://www.wired.com/2016/03/final-game-alphago-lee-sedol-big-deal-humanity/

Minsky M. & Papert S. (1969) *Perceptrons.* Cambridge, MA: MIT Press.

Naisbitt J. & Aburdene P. (1990) *Megatrends 2000. Ten New Directions for the 1990's.* New York: William Morrow.

Noga H. (2016) Selected Factors of the Socialization Impact of Computer Games and Their Content. *Society. Integration. Education: Proceedings of the International Scientific Conference.* Vol. 2, pp. 569–578.

Novikov F.A. (2019) *Symbolic Artificial Intelligence: Mathematical Foundations of Knowledge Representation.* Moscow: Urait (in Russian).

Penrose R. (1989) *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford: Oxford University Press (Russian translation: Moscow: URSS, Lenand, 2020).

Perse E.M. & Lambe J. (2016) *Media Effects and Society.* New York: Routledge.

Rosenblatt F. (1961) *Principles of Neurodynamics.* Washington, DC: Spartan Books.

Rumelhart D.E., Hinton G.E., & Williams R.J. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations.* Cambridge, MA: MIT Press.

Santos H.C., Varnum M.E.W., & Grossmann I. (2017) Global Increases in Individualism. *Psychological Science.* Vol. 28, no. 9, pp. 1228–1239.

Searle J. (2001) Chinese Room Argument. In: Wilson R.A. & Keil F.C. (Eds.) *The MIT Encyclopedia of the Cognitive Science* (pp. 115–116). Cambridge, MA: MIT Press.

Silver D., Huang A., Maddison C. et al. (2016) Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature.* Vol. 529, no. 7587, pp. 484–489.

Sokolenko O. (2020, September 25) "Today, the general level of AI is comparable to the level of intelligence of a mouse or bee. Are we afraid that mice and bees will want to take over the world?" Head of MIPT Laboratory of Neural Systems and Deep Learning Mikhail Burtsev – on the risks of artificial intelligence for humanity. *Republic.* Retrieved from: https://republic.ru/posts/97895 (in Russian).

Varlamov O.O. (2011) Mivar's Approach to Development of Intelligent Systems and Project of Development of Active Multi-Subject Online Mivar's Internet Encyclopedia. *Bulletin of Kabardino-Balkaria Scientific Center of RAS.* No. 1, pp. 55–64 (in Russian).

Walter W.G. (1950) An Imitation of Life. *Scientific American.* Vol. 182, no. 5, pp. 42–45.

Weizenbaum J. (1966) ELIZA – a Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM.* Vol. 9, no. 1, pp. 36–45.

НАУЧНАЯ ЖИЗНЬ

Обзоры, объявления, сообщения

DOI: 10.30727/0235-1188-2021-64-1-149-154
Обзор мероприятий
Events summary

Санкт-Петербургское отделение Научного совета по методологии искусственного интеллекта и когнитивных исследований РАН в 2020 году

С.Ф. Сергеев
Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

Аннотация

В обзоре рассмотрены история создания, состав и направления деятельности Санкт-Петербургского отделения Научного совета по методологии искусственного интеллекта и когнитивных исследований при Президиуме РАН. Члены отделения ориентированы на практику, что обусловлено преобладанием в его составе ученых инженерного и естественно-научного профиля. Совет, согласно авторской позиции, стимулирует работы, связанные с междисциплинарным синтезом и конвергенцией гуманитарных и естественных наук при решении проблемы искусственного интеллекта.

Ключевые слова: методология и теория искусственного интеллекта, философия искусственного интеллекта, когнитивные науки, проблемы междисциплинарного синтеза, конвергенция гуманитарных и естественных наук.

Сергеев Сергей Федорович – доктор психологических наук, профессор кафедры информационных систем в искусстве и гуманитарных науках Санкт-Петербургского государственного университета, заведующий научно-исследовательской лабораторией «Эргономика сложных систем» Санкт-Петербургского политехнического университета Петра Великого.

s.f.sergeev@spbu.ru
http://orcid.org/0000-0002-6677-8320

Для цитирования: *Сергеев С.Ф.* Санкт-Петербургское отделение Научного совета по методологии искусственного интеллекта и когнитивных исследований РАН в 2020 году // Философские науки. 2021. Т. 64. № 1. С. 149–154. DOI: 10.30727/0235-1188-2021-64-1-149-154

Saint Petersburg Branch of the RAS Scientific Council on the Methodology of Artificial Intelligence and Cognitive Research in 2020

S.F. Sergeev
Saint Petersburg State University, Saint Petersburg, Russia;
Peter the Great Saint Petersburg Polytechnic University,
Saint Petersburg, Russia

Abstract

The summary discusses the history, structure, and areas of activity of the Saint Petersburg branch of the Scientific Council on the Methodology of Artificial Intelligence and Cognitive Research under the Presidium of the Russian Academy of Sciences. The members of the Saint Petersburg branch are focused on practice, which is due to the predominance of researchers in the engineering and natural sciences. According to the author, the Council stimulates work related to interdisciplinary synthesis and convergence of the humanities and natural sciences in solving the problem of artificial intelligence.

Keywords: methodology and theory of artificial intelligence, philosophy of artificial intelligence, cognitive sciences, problems of interdisciplinary synthesis, convergence of the humanities and natural sciences.

Sergey F. Sergeev – D.Sc. in Psychology, Professor, Department of Information Systems in the Arts and Humanities of Saint Petersburg State University; Head of the Research Laboratory "Ergonomics of Complex Systems," Peter the Great Saint Petersburg Polytechnic University.
s.f.sergeev@spbu.ru
http://orcid.org/0000-0002-6677-8320

For citation: Sergeev S.F. (2021) Saint Petersburg Branch of the RAS Scientific Council on the Methodology of Artificial Intelligence and Cognitive Research in 2020. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 149–154.
DOI: 10.30727/0235-1188-2021-64-1-149-154

В состав Санкт-Петербургского отделения Научного совета по методологии искусственного интеллекта и когнитивных исследований (НСМИИ и КИ) при Президиуме Российской академии наук вошли действующие ученые, работающие в сфере методологии, теории и практики технологий искусственного интеллекта. Они представляют широкий спектр научных организаций, находящихся в структуре РАН, высшего образования, науки и промышленности Петербурга и Северо-Западного региона.

В Санкт-Петербургское отделение Совета избраны представители Санкт-Петербургского политехнического университета Петра Великого (СПбПУ) – профессора В.С. Заборовский (д. техн. н.), Л.С. Уткин (д. техн. н.), С.Ф. Сергеев (д. психол. н.), Л.А. Станкевич (д. техн. н.), О.С. Ипатов (д. техн. н.), С.В. Козырев (канд. техн. н.); Санкт-Петербургского института информатики и автоматизации РАН – А.Л. Ронжин (проф., д. техн. н.), А.А. Карпов (доц., д. техн. н.); Института эволюционной физиологии и биохимии имени И.М. Сеченова РАН – М.Л. Фирсов (д. биол. н.), Н.В. Шемякина (канд. биол. н.), Е.И. Гальперина (канд. биол. н.); Санкт-Петербургского государственного университета (СПбГУ) – профессора С.Н. Костромина (д. психол. н.), О.Н. Граничин (д. физ.-мат. н.); Института физиологии имени И.П. Павлова РАН – Е.А. Рыбникова (д. биол. н.), Ю.Е. Шелепин (д. мед. н.), Р.О. Малашин (канд. техн. н.); Центрального научно-исследовательского института робототехники и технической кибернетики – Е.И. Юревич (проф., д. техн. н.), А.Б. Железняков (канд. техн. н.); Национального медицинского исследовательского центра имени В.А. Алмазова – чл.-корр. РАН А.О. Конради (проф., д. мед. н.).

Председателем Санкт-Петербургского отделения НСМИИ и КИ РАН избран ***С.Ф. Сергеев***, ученым секретарем – ***С.В. Козырев***.

Деятельность Санкт-Петербургского отделения НСМИИ и КИ в 2020 году

Члены регионального отделения проявляют активность и интерес к научным мероприятиям, которые проводит Совет, формируют их повестку. Так, по инициативе члена отделения, проф. О.Н. Граничина под эгидой НСМИИ и КИ РАН 30 апреля 2020 года в режиме *Zoom*-конференции состоялся круглый стол «Самоорганизация и искусственный интеллект в группе автономных роботов: методология, теория, практика», где участвовали более 50 специалистов в области информатики, робототехники,

технологий искусственного интеллекта, инженерной психологии и эргономики.

В работе круглого стола приняли активное участие представители СПбГУ – советник проректора по научной работе, проф. кафедры информатики А.Л. Тулупьев, директор научно-образовательного центра «Математическая робототехника и искусственный интеллект» К.С. Амелин, заведующий кафедрой математического моделирования энергетических систем, проф. В.В. Захаров, заведующий кафедрой системного программирования А.Н. Терехов и доц. кафедры С.С. Сысоев, доц. кафедры арабской филологии, сотрудник научной лаборатории анализа и моделирования социальных процессов О.А. Берникова, доц. биологического факультета М.П. Иванов, старший преподаватель кафедры информатики С.И. Салищев, студенты-магистры отделения программной инженерии И.А. Лень и А.Н. Сергеенко. Участниками мероприятия стали представители АО «Концерн "ЦНИИ "Электроприбор"» (Санкт-Петербург) – чл.-корр. РАН, начальник научно-образовательного центра Концерна, профессор Санкт-Петербургского национально-исследовательского университета информационных технологий, механики и оптики (Университета ИТМО) О.А. Степанов, начальник сектора Концерна, доц. Университета ИТМО А.В. Лопарев, ведущий научный сотрудник Концерна А.Е. Пелевин, младший научный сотрудник Концерна О.В. Зайцев. Активно участвовали в работе круглого стола проф. кафедры электронных систем и информационной безопасности Самарского государственного технического университета П.О. Скобелев, заведующий научно-исследовательской лабораторией математического моделирования Ульяновского государственного педагогического университета имени И.Н. Ульянова, проф. кафедры высшей математики данного университета А.В. Цыганов, проф. кафедры информационных технологий Ульяновского государственного университета Ю.В. Цыганова, руководитель магистерского отделения факультета программной инженерии колледжа ОРТ имени Брауде (Израиль), проф. В. Волькович и старший преподаватель, координатор студенческих проектов колледжа Р. Аврос.

Главная цель круглого стола – обсуждение проблем возникновения в сложных интеллектуальных системах самоорганизующихся структур за счет коммуникации и многоуровневой кооперации между компонентами системы. В круг обсуждавшихся проблем

вошли не только вопросы организации движения группы роботов, формирования «эмерджентного» интеллекта, исследований интеллектуального поведения насекомых и животных, но и возможностей создания симбиотических форм искусственного интеллекта и искусственного разума, а также отдельных аспектов работы психики и мозга человека в циклах социальной самоорганизации техногенных сред. По итогам докладов, с которыми выступили участники круглого стола, и последовавшей за ними дискуссии издана коллективная монография «Самоорганизация и искусственный интеллект в группах автономных роботов: методология, теория, практика» [Самоорганизация… 2020].

19 ноября 2020 года в формате *Zoom*-конференции, объединившей свыше 150 человек, прошло совместное заседание научно-практического семинара «Актуальные проблемы психологии труда, инженерной психологии и эргономики», организованного Институтом психологии РАН совместно с Санкт-Петербургским отделением НСМИИ и КИ при Президиуме РАН. Тема заседания – «Искусственный интеллект: вызовы для психологии и эргономики».

В рамках программы семинара, после приветственного слова директора Института психологии РАН, академика РАН Д.В. Ушакова, заслушаны доклады председателя Санкт-Петербургского отделения НСМИИ и КИ при Президиуме РАН, проф. СПбГУ, заведующего лабораторией «Эргономика сложных систем» СПбПУ С.Ф. Сергеева («Психологические проблемы искусственного интеллекта»); заместителя генерального директора по гражданской авиации Государственного научно-исследовательского института авиационных систем, проф. РАН, д. техн. н. В.В. Косьянчука («Проблемы интеллектуализации человеко-машинного интерфейса авиационных комплексов»); президента Российской ассоциации нейроинформатики, директора научно-координационного совета Центра компетенций Национальной технологической инициативы по направлению «Искусственный интеллект» Московского физико-технического института, канд. физ.-мат. н. С.А. Шумского («Проблемы создания искусственной психики»); заместителя заведующего кафедрой робототехнических систем и мехатроники Московского государственного технического университета имени Н.Э. Баумана, члена Научного совета по робототехнике при Президиуме РАН, д. техн. н., проф. А.С. Ющенко («Психологические проблемы коллаборативной робототехники»);

заведующего лабораторией социальной психологии Института психологии РАН, проф. РАН, д. психол. н., проф. Т.А. Нестика («Социально-психологические предпосылки отношения личности к использованию систем искусственного интеллекта в период пандемии COVID-19»).

Далее состоялась дискуссия с участием ряда ученых. Среди них – профессора Н.Б. Филимонов (д. техн. н.), А.А. Обознов (д. психол. н.), чл.-корр. РАО В.И. Панов (д. психол. н.), А.А. Грачёв (д. психол. н.), И.Г. Городецкий (канд. техн. н.), В.В. Толочек (д. психол. н.), П.И. Падерно (д. техн. н.), С.Н. Костромина (д. психол. н.), О.Н. Граничин (д. физ.-мат. н.); ведущий научный сотрудник Института психологии РАН И.Г. Скотникова (д. психол. н.), ученые М.В. Найченко (д. техн. н.), Е.А. Трофимов, С.А. Шумский (канд. физ.-мат. н.).

Анализ выступлений показывает ярко выраженную направленность работы Санкт-Петербургского отделения на решение задач междисциплинарного синтеза областей теоретического и методологического знания философских, гуманитарных и инженерно-технологических дисциплин в форме конвергентных образований с гибкими динамическими границами.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Самоорганизация… 2020 – Самоорганизация и искусственный интеллект в группах автономных роботов: методология, теория, практика: коллективная монография / под ред. О.Н. Граничина, С.Ф. Сергеева. – СПб.: Изд-во ВВМ, 2020.

REFERENCES

Granichin O.N. & Sergeev S.F. (Ed.) (2020) *Self-Organization and Artificial Intelligence in Groups of Autonomous Robots: Methodology, Theory, Practice*. Saint Petersburg: VVM Publishing (in Russian).

DOI: 10.30727/0235-1188-2021-64-1-155-159
Обзор мероприятий
Events summary

О деятельности Самарского отделения Научного совета по методологии искусственного интеллекта и когнитивных исследований РАН

А.Ю. Нестеров, А.В. Никоноров, А.В. Куприянов
Самарский национальный исследовательский университет имени академика С.П. Королева, Самара, Россия

Аннотация

В обзоре изложены основные результаты деятельности Самарского отделения Научного совета по методологии искусственного интеллекта и когнитивных исследований РАН, созданного в 2007 году на базе Самарского национального исследовательского университета имени академика С.П. Королева. Самарским отделением НСМИИ и КИ РАН совместно с Самарским университетом проведены международные конференции по информационным технологиям, информационному обществу, научной фантастике, создан Центр искусственного интеллекта, реализованы междисциплинарные технико-гуманитарные исследовательские проекты в сфере социогуманитарной кибернетики, цифровых моделей творческих процессов, вычислительной эстетики.

Ключевые слова: методология и теория искусственного интеллекта, философия искусственного интеллекта, философия техники, философия кибернетики, футурология.

Нестеров Александр Юрьевич – доктор философских наук, доцент, директор Социально-гуманитарного института Самарского национального исследовательского университета имени академика С.П. Королева (Самарского университета), заведующий кафедрой философии Самарского университета.

phil@ssau.ru
https://orcid.org/0000-0002-0670-9315

Никоноров Артем Владимирович – доктор технических наук, директор Института искусственного интеллекта Самарского университета.

artniko@gmail.com
https://orcid.org/0000-0003-4292-2049

Куприянов Александр Викторович – доктор технических наук, доцент, заведующий кафедрой технической кибернетики Самарского университета.

akupr@ssau.ru

https://orcid.org/0000-0002-0436-4392

Для цитирования: *Нестеров А.Ю., Никоноров А.В., Куприянов А.В.* О деятельности Самарского отделения Научного совета по методологии искусственного интеллекта и когнитивных исследований РАН // Философские науки. 2021. Т. 64. № 1. С. 155–159.
DOI: 10.30727/0235-1188-2021-64-1-155-159

On the Work of the Samara Branch of the RAS Scientific Council on the Methodology of Artificial Intelligence and Cognitive Research

A.Yu. Nesterov, A.V. Nikonorov, A.V. Kupriyanov
S.P. Korolev Samara National Research University, Samara, Russia

Abstract

The summary presents the main results of the work of the Samara branch of the RAS Scientific Council for the Methodology of Artificial Intelligence and Cognitive Research, created in 2007 on the basis of S.P. Korolev Samara National Research University (Samara University). The Samara branch of the Council and the Samara University held international conferences on information technology, information society, science fiction, established Artificial Intelligence Center as well as completed interdisciplinary technical and humanitarian research projects in the field of socio-humanitarian cybernetics, digital models of creative processes, computational aesthetics.

Keywords: methodology and theory of artificial intelligence, philosophy of artificial intelligence, philosophy of technology, philosophy of cybernetics, futurology.

Alexander Yu. Nesterov – D.Sc. in Philosophy, Associate Professor, Director of the Institute of Social Sciences and Humanities, Samara University, Head of the Department of Philosophy, Samara University.

phil@ssau.ru

https://orcid.org/0000-0002-0670-9315

Artem V. Nikonorov – D.Sc. in Technology, Director of the Institute of Artificial Intelligence, Samara University.

artniko@gmail.com

https://orcid.org/0000-0003-4292-2049

Alexander V. Kupriyanov – D.Sc. in Technology, Associate Professor, Head of the Department of Technical Cybernetics, Samara University.
akupr@ssau.ru
https://orcid.org/0000-0002-0436-4392

For citation: Nesterov A.Yu., Nikonorov A.V., & Kupriyanov A.V. (2021) On the Work of the Samara Branch of the RAS Scientific Council on the Methodology of Artificial Intelligence and Cognitive Research. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 155–159.
DOI: 10.30727/0235-1188-2021-64-1-155-159

Самарское региональное отделение Научного совета по методологии искусственного интеллекта Российской академии наук (НСМИИ РАН) на базе факультета информатики и кафедры философии Самарского государственного аэрокосмического университета имени академика С.П. Королева (национального исследовательского университета) утверждено НСМИИ РАН на заседании, состоявшемся 26 сентября 2007 года. Руководителем Самарского регионального отделения является доктор технических наук, профессор ***С.А. Прохоров***, заместителем – доктор философских наук, профессор ***Р.И. Таллер***, ученым секретарем – доктор философских наук, доцент ***А.Ю. Нестеров***.

В настоящее время отделение функционирует на базе Самарского национального исследовательского университета имени академика С.П. Королева (Самарского университета). Отделение организует две регулярных международных конференции, которые ориентированы на технические и гуманитарные аспекты проблематики искусственного интеллекта. Ежегодно факультет информатики выступает организатором конференции «Перспективные информационные технологии», включающей в себя гуманитарную секцию «Философия искусственного интеллекта и трансгуманизм» [Перспективные информационные технологии… 2020]. Каждые два года кафедра философии проводит «Лемовские чтения» [Пятые Лемовские чтения… 2020], единственную в мире регулярную научную конференцию, посвященную научной фантастике, где особое внимание уделено обсуждению широкого спектра вопросов в сферах философии техники и футурологии.

С 2015 года на базе Самарского университета ежегодно проводится Международная конференция и молодежная школа

«Информационные технологии и нанотехнологии» (ИТНТ)[1]. Цель Конференции ИТНТ – предоставление возможности научных дискуссий, обсуждения результатов фундаментальных и прикладных исследований в области нанотехнологий и информационных технологий. Рабочими языками Конференции являются английский и русский. В рамках Конференции работает секция «Data Science», тематика которой в первую очередь включает в себя технологии искусственного интеллекта для анализа данных (нейронные сети, глубокое обучение, большие данные), а также программные средства для искусственного интеллекта (платформы, библиотеки, аппаратные средства, параллельные алгоритмы, облачные вычисления). В 2020 году в условиях пандемии все конференции состоялись в смешанном формате. По итогам изданы сборники трудов, которые доступны в открытых источниках. Издания Конференции *ITNT* 2020 индексируются в РИНЦ, *Scopus* и *Web Of Science*.

За 13 лет успешной работы десятки участников регионального отделения защитили кандидатские и докторские диссертации. В 2020 году, учитывая резкое повышение интереса в мире к системам искусственного интеллекта, к их правовому сопровождению, необходимость социогуманитарной экспертизы и всестороннего рефлексивного анализа ситуаций, вызываемых к жизни внедрением этих систем, в Самарском университете принят ряд решений, значимых для работы отделения. Создан Центр искусственного интеллекта[2], открыты новые междисциплинарные программы магистерской подготовки «Философия кибернетики» и «Цифровая педагогика и киберпсихология», разработан социогуманитарный профиль подготовки бакалавров направления «Прикладная математика и информатика».

В сфере технической кибернетики компетенции Самарского университета в области искусственного интеллекта определяются научной школой академика В.А. Сойфера. На базе кафедры технической кибернетики функционируют с 2014 года – лаборатория обработки данных сверхбольшого объема (*Big Data Lab*), с 2018 года –

[1] Сайт Конференции ИТНТ: http://itnt-conf.org/

[2] Работа Центра искусственного интеллекта освещается в интервью директора А.В. Никонорова сетевому изданию «Интерфакс» (см.: *Никоноров А.В.* Мы хотим научить искусственный интеллект чувствовать красоту // Интерфакс. 2020. 9 декабря. – URL: https://www.interfax-russia.ru/volga/exclusives/direktor-centra-iskusstvennogo-intellekta-samarskogo-universiteta-artem-nikonorov-my-hotim-nauchit-robota-chuvstvovat-krasotu).

лаборатория систем искусственного интеллекта, с 2019 года – лаборатория виртуальной и дополненной реальности. В рамках направления 01.04.02 «Прикладная математика и информатика» с 2018 года совместно с ПАО «Сбербанк» реализуется магистерская программа «Науки о данных (*Data Science*)»[3].

В сфере междисциплинарных технико-гуманитарных исследований в университете запущен ряд новых научных проектов, связанных с социогуманитарной кибернетикой, разработкой цифровых моделей творческих процессов, с вычислительной эстетикой. Новое планируемое регулярное мероприятие – ежегодная Международная научная конференция «Человек в информационном обществе», ориентированная на анализ, интерпретацию и создание прогнозных сценариев влияния технологий искусственного интеллекта на человека, его быт, способы и формы рефлексии[4].

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Перспективные информационные технологии… 2020 – Перспективные информационные технологии: сб. науч. тр. Междунар. науч.-техн. конф. / под ред. С.А. Прохорова. – Самара: Изд-во Самарского научного центра РАН, 2020.

Пятые Лемовские чтения… 2020 – Пятые Лемовские чтения: сб. матер. Междунар. науч. конф. памяти Станислава Лема / отв. ред. А.Ю. Нестеров. – Самара: Самарская гуманитарная академия, 2020.

REFERENCES

Prokhorov S.A. (Ed.) (2020) *Perspective Information Technologies: Proceedings of the International Scientific-Technical Conference.* Samara: RAS Samara Scientific Center Publishing (in Russian).

Nesterov A.Yu. (Ed.) (2020) *The Fifth Lem Readings: Proceedings of Stanislav Lem International Scientific Conference.* Samara: Samara Academy of Humanities (in Russian).

[3] Более подробная информация о магистерской программе доступна на интернет-странице: https://sbergraduate.ru/program/2715/

[4] Выступления участников конференции доступны на youtube-канале Самарского университета. (см., например: *Сойфер В.А.* Эволюция общества: от информационного 4.0 к суперинтеллектуальному 5.0. – URL: https://www.youtube.com/watch?v=O6-qLOkWgUQ).

Цели и задачи

«Философские науки» – ежемесячный рецензируемый научный журнал, публикующий статьи на русском и английском языках. Журнал был учрежден в 1958 г. как научное образовательное просветительское издание для системы отечественного образования. Статьи журнала посвящены как традиционным, классическим философским темам, так и актуальным проблемам современности и перспективам социокультурного и цивилизационного развития человечества.

Aims and Scope

Russian Journal of Philosophical Sciences = Filosofskie nauki is a peer-reviewed monthly journal published in Russian and English languages. The journal was founded in 1958 as a scientific and educational periodical for the educational system. The papers of the journal are dedicated to classical problems of philosophy as well as to important issues of the modernity and prospects of sociocultural and civilizational development of humankind.

С условиями публикации материалов в журнале
и порядком рецензирования статей можно ознакомиться на сайте:
https://www.phisci.info/jour/about/submissions#authorGuidelines

Подписной индекс в Объединенном каталоге «Пресса Россия» – 45490
Журнал зарегистрирован в Министерстве РФ по делам печати, телерадиовещания и средств массовых коммуникаций. Свидетельство о регистрации ПИ № 77 – 15513 от 20 мая 2003 г.

Подписано в печать 15.01.2021 г. Формат 60x90/16. Печать цифровая.
Бумага офсетная № 1. Печ. л. 10,0. Тираж 1000 экз. Заказ
Отпечатано в типографии Издательского дома «Гуманитарий».
E-mail: humanist@academyrh.info