ПРОЕКТ НАЦИОНАЛЬНОГО КОДЕКСА ЭТИКИ В СФЕРЕ ИИ

ПАО Сбербанк / Альянс ИИ, А. Незнамов
Октябрь 2021

ЭТИКА В СФЕРЕ ИИ: НОВЫЕ ЗАКОНЫ РОБОТОТЕХНИКИ

ОТ «РОБОЭТИКИ» → К ЭТИКЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Лига Гуманности: роботы не должны подвергаться бесчеловечному обращению (R.U.R. Карела Чапека, 1920).

Четыре закона робототехники (Айзек Азимов, 1942)

Robot Ethics Charter (Южная Корея, 2007)

Европейская хартия робототехники (Европарламент, 2017)

10 Законов для искусственного интеллекта (Microsoft, 2016)

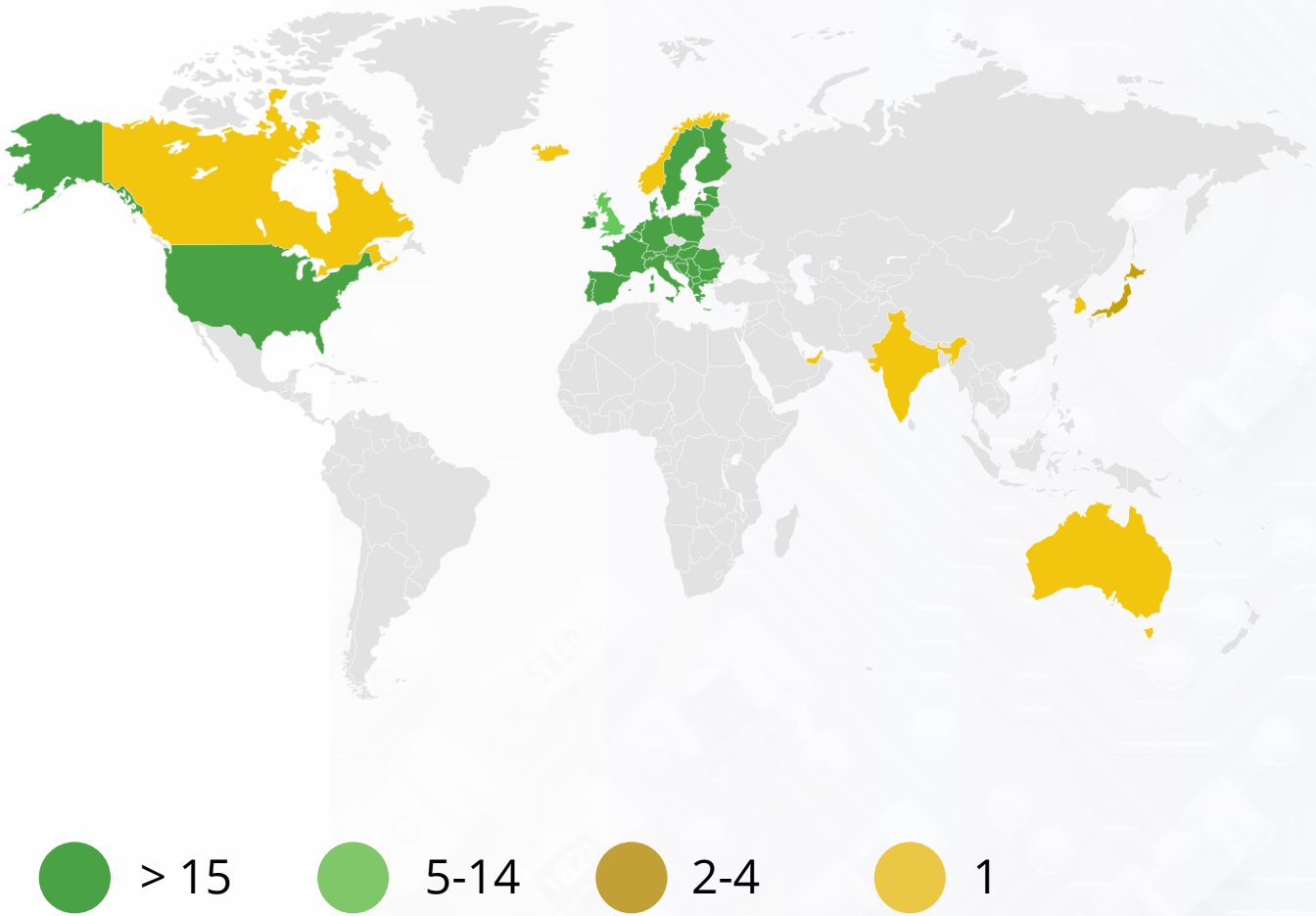
Азиломарские принципы искусственного интеллекта (Future of life Institute, 2017)

Глобальная инициатива по этике автономных и интеллектуальных систем (IEEE, 2017)

Декларация о сотрудничестве в сфере искусственного интеллекта (страны ЕС, 2018)

Принципы искусственного интеллекта (ОЭСР, 2019)

ЭТИКА В СФЕРЕ ИИ: СТАТИСТИКА



> 100

самостоятельных этических документов на всех уровнях

> 1100

этических документов, связанных с ИИ

> 20

стран занимаются этикой ИИ

> 5000

исследователей в сфере этики ИИ

< 10

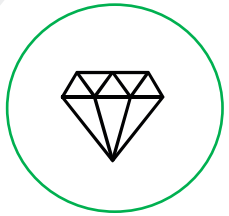
ключевых принципов ИИ

Все ключевые международные организации имеют инициативы в сфере этики ИИ

The global landscape of AI ethics guidelines by Anna Jobin, Marcello Ienca and Effy Vayena

Наиболее распространенные принципы ИИ в мире

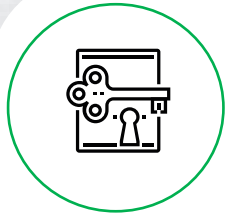
Справедливость, включая
- недискриминацию
- непредвзятость
- инклюзивность
- равенство
- равный доступ к благам ИИ
- минимизацию негативных последствий на рынке труда



Прозрачность, включая
- прозрачность использования данных и дата сетов
- прозрачность обработки информации
- прозрачность во взаимодействии человека с ИИ
- прозрачность алгоритмического принятия решений

Ответственность, включая
- контролируемость
- подотчетность
- возмещение вреда
- ответственное отношение

Безопасность, включая
- невозможность причинения умышленного вреда



Конфиденциальность, включая
- неприкосновенность частной жизни
- гарантии всех базовых прав
- защиту персональных данных

ПРЕДЛОЖЕНИЯ НКО И ЧАСТНЫХ ГРУПП ИССЛЕДОВАТЕЛЕЙ

OpenAI

< >
Déclaration de Montréal
IA responsable_
</ >

AINOW

23 Asilomar AI principles

Montréal Declaration: Responsible AI

The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term

OpenAI Charter

The Toronto Declaration: Protecting the Right to Equality and Non-discrimination in Machine Learning System

Группа проф. Susumu Hirano, Япония: 8 принципов

Модельная конвенция робототехники и ИИ (Россия)

КОРПОРАТИВНЫЕ НОРМЫ

«10 Законов для искусственного интеллекта», Microsoft

«7 принципов ИИ», Google

«6 тематических областей исследования этики ИИ», Deepmind

«5 принципов ИИ», Telefonica

«Руководство по ИИ», Deutsche Telekom

«Руководящие принципы ИИ», SAP

«Этические принципы», Японская ассоциация ИИ

«Руководящие принципы этики ИИ», Sony Group

«Руководящие принципы ИИ», Unity

«Инициатива в сфере ИИ», Partnership on AI

ЦИОНАЛЬНЫЕ ИНИЦИАТИВЫ

- Robot Ethics Charter (Южная Корея)
- A guide to using artificial intelligence in the public sector (Великобритания)
- Understanding artificial intelligence ethics and safety (Великобритания)
- AI Ethics Framework (Австралия)
- Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government (США)
- Ethical Principles for Artificial Intelligence (США)
- Principles of Artificial Intelligence Ethics for the Intelligence Community (США)
- Civil Law Rules on Robotics (European Parliament)
- Руководящие принципы применения ИИ (Канада)
- Social principles of Human-centric AI (Япония)
- Artificial intelligence at the service of citizens (Италия)
- AI ethics and governance body of knowledge (Сингапур)
- Ethics guidelines for intelligent artificial society (Южная Корея)
- AI Principles & Ethics (Smart Dubai)

НАДНАЦИОНАЛЬНЫЕ ИНИЦИАТИВЫ

Проект рекомендаций по этике ИИ ЮНЕСКО

- Ethically Aligned Design (IEEE)
- Top 10 Principles for Ethical Artificial Intelligence (UNI Global Union)
- Ethics Guidelines for Trustworthy AI (Совет Европы)
- European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment (Совет Европы)
- Report on Robotics Ethics (COMEST OOH)
- Principles on AI (ОЭСР)

ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ В РФ

Национальная стратегия развития искусственного интеллекта до 2030 г., утверждена Указом Президента Российской Федерации от 10 октября 2019 г. № 490:
«48. Для стимулирования развития и использования технологий искусственного интеллекта необходимы адаптация нормативного регулирования в части, касающейся взаимодействия человека с искусственным интеллектом, и выработка соответствующих этических норм.».

Федеральный проект «Искусственный интеллект».

Стратегия развития информационного общества в Российской Федерации на 2017 – 2030 годы

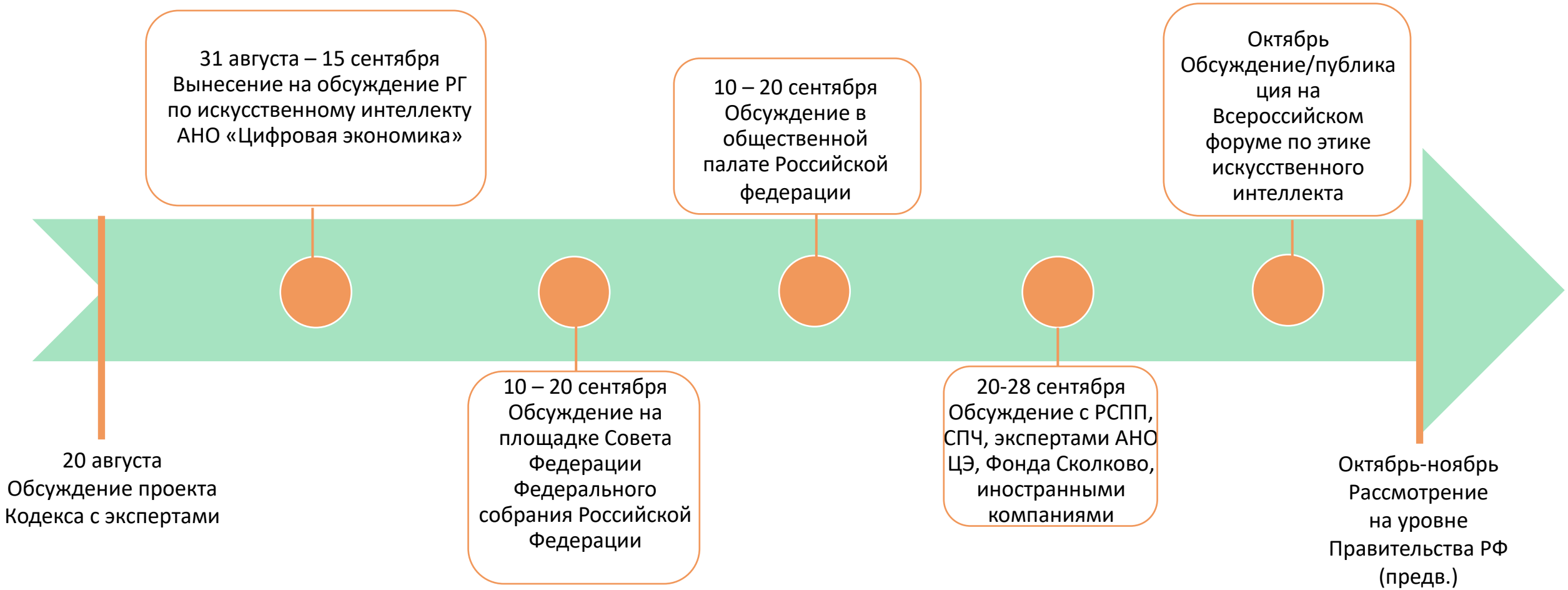
Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года, утверждена Распоряжением Правительства Российской Федерации от 19 августа 2020 г. № 2129-р.:
«Следует поддерживать развитие регулирования, вырабатываемого и приводимого в исполнение силами участников рынка (саморегулирование), включая принятие и использование документов национальной системы стандартизации, кодексов (сводов) этических правил и иных документов саморегулируемых организаций, а также иных инструментов.».

> *Кроме различных административных, законодательных ограничений, мне кажется, нужно подумать, я в прошлом году тоже об этом сказал, о том,* ***чтобы выработать для этой среды такой внутренний морально-нравственный кодекс работы****.*

В.В. Путин, Конференция по искусственному интеллекту AIJ, 20.12.2020 г.

РАЗРАБОТКА НАЦИОНАЛЬНОГО КОДЕКСА

Проект рекомендаций по этике ИИ разработан на площадке Альянса в сфере ИИ (Газпромнефть, МТС, Мэйл.Ру, РФПИ, Сбер, Яндекс) совестно с представителями государства и при участии представителей научного сообщества



НАЦИОНАЛЬНЫЙ КОДЕКС ЭТИКИ ИИ

Общая характеристика Кодекса:

- Первый документ подобного рода в России
- Носит рекомендательный характер
- Присоединение осуществляется на добровольной основе
- Устанавливает общие этические принципы и стандарты поведения, для российских и иностранных участников отношений в сфере ИИ (акторов ИИ)
- Устанавливает механизмы реализации положений Кодекса
- Действие Кодекса распространяется на системы ИИ, применяемые исключительно в гражданских (не военных) целях

Механизм реализации:

- Акторам ИИ рекомендуется назначить Уполномоченного по этике, ответственного за реализацию Кодекса
- Акторы ИИ могут создавать Комиссии по этике в сфере ИИ
- Акторы ИИ могут создавать публичный свод наилучших и/или наихудших практик решения возникающих этических вопросов в жизненном цикле ИИ
- Рекомендуется разработка методик, обеспечивающих соблюдение положений Кодекса
- На сайте Комиссии по этике ИИ ведется публичный Реестр Акторов ИИ
- Появится Национальная комиссия по этике ИИ на площадке Альянса в сфере ИИ

СОДЕРЖАНИЕ НАЦИОНАЛЬНОГО КОДЕКСА ЭТИКИ ИИ (1/2)

Содержание Кодекса базируется на 7 принципах, положенных в основу детальных рекомендаций:

1

Главный приоритет развития технологий ИИ – защита интересов людей, отдельных групп, каждого человека.

Реализация принципа заключается в:

- человеко-ориентированности технологий
- полном соответствии закону
- недискриминации
- проведении оценки рисков и гуманитарного воздействия в необход

2

Необходимость осознания ответственности при создании и использовании ИИ.

Реализация принципа заключается в:

- применении риск-ориентированного подхода
- ответственном отношении к влиянию ИИ на общество и граждан
- предосторожности и непричинении вреда
- идентификации ИИ в общении с человеком и уважении его автономии воли
- безопасности работы с данными и информационной безопасности

и соответствии положениям

3

Ответственность за последствия применения ИИ всегда лежит на человеке.

ания систем ИИ (создании

Реализация принципа заключается в:

- поднадзорности ИИ
- ответственности за последствия принятия решений ИИ

СОДЕРЖАНИЕ НАЦИОНАЛЬНОГО КОДЕКСА ЭТИКИ ИИ (2/2)

4

Технологии ИИ внедрять там, где это принесёт пользу людям.

Реализация принципа заключается в:

- применении ИИ в соответствии с предназначением
- стимулировании развития ИИ

5

Интересы развития технологий ИИ выше интересов конкуренции.

Реализация принципа заключается в:

- корректности сравнения систем ИИ
- развитии профессиональных компетенций в области ИИ
- сотрудничестве разработчиков ИИ

6

Важна максимальная прозрачность и правдивость в информировании об уровне развитии технологий ИИ, их возможностях и рисках.

Реализация принципа заключается в:

- достоверности информации о системах ИИ, предоставляемой пользователям
- повышении осведомлённости граждан и общества об этике применения ИИ

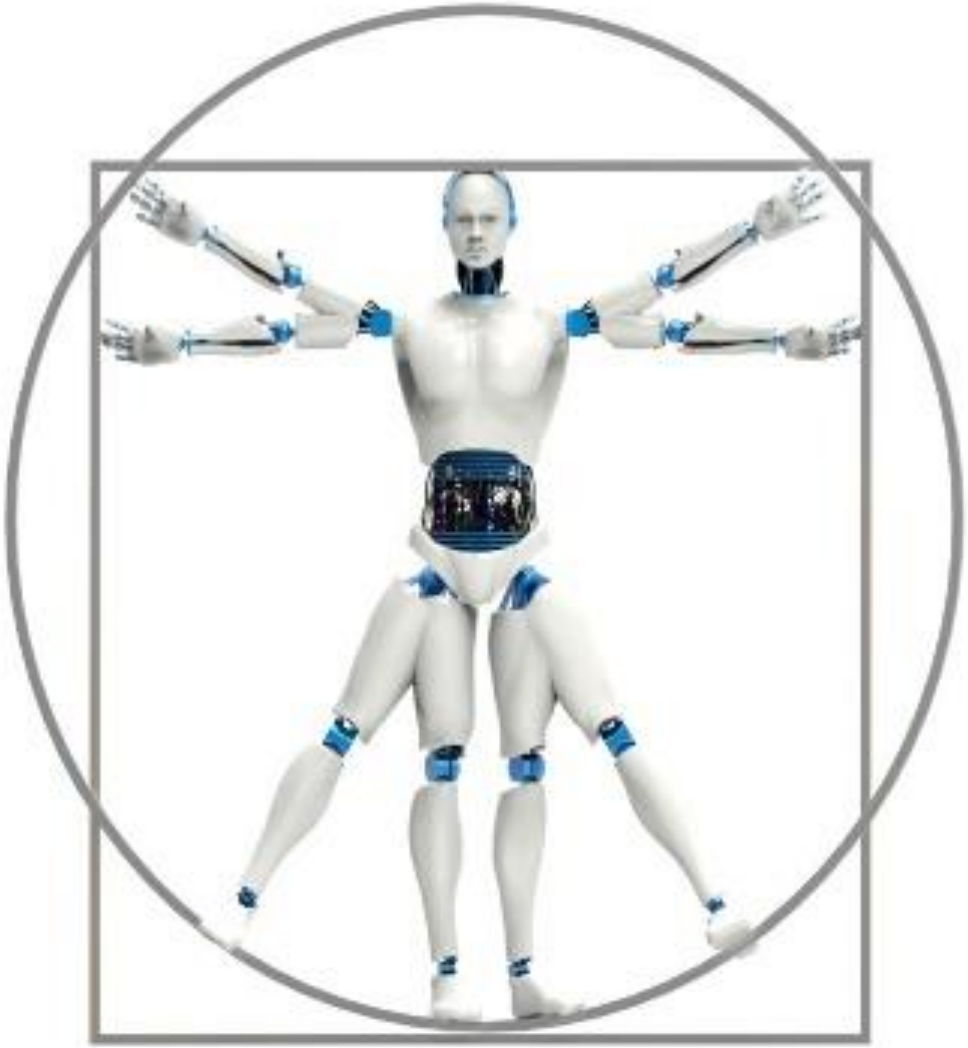
7

Принципы этики развиваются по мере появления новых знаний, новых вызовов, новых возможностей.

Реализация принципа заключается в:

- развитии принципов соразмерно с развитием технологий ИИ

Спасибо
за внимание!



Neznamov.A.V@sberbank.r
u