

ГАРДА

Безопасность ИИ как кросс-дисциплинарный вопрос

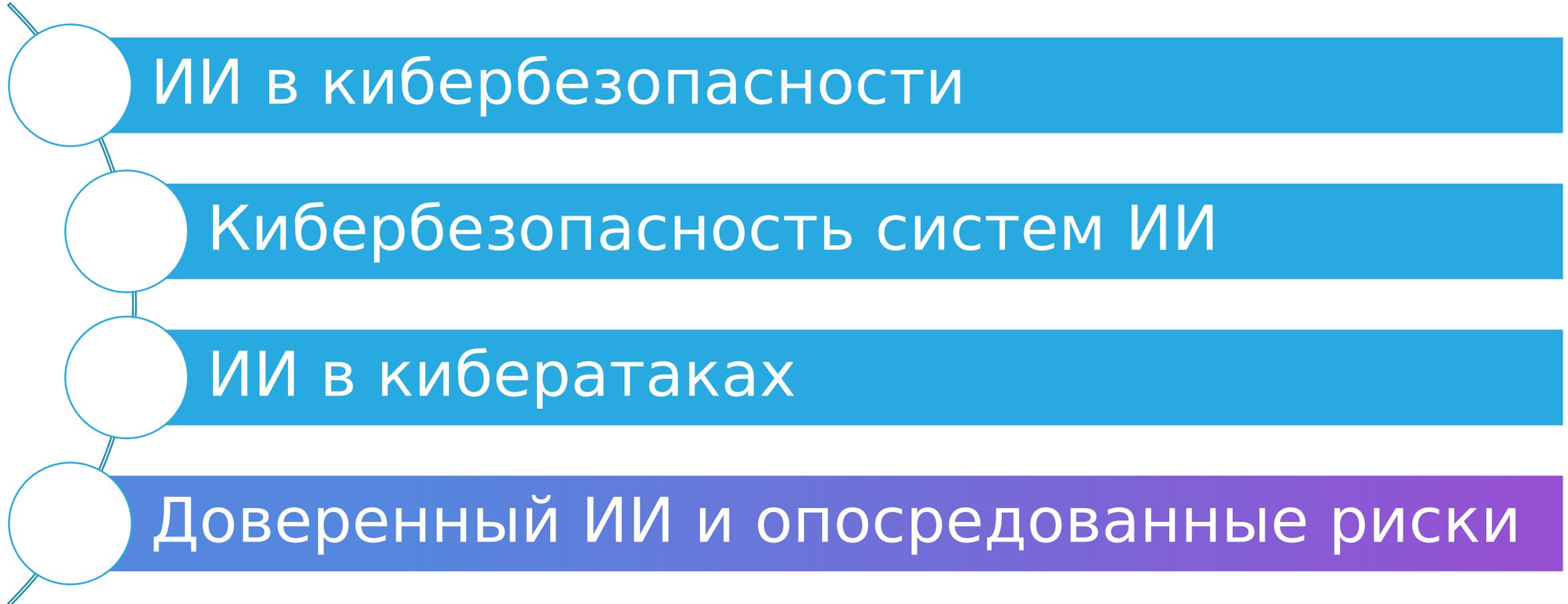
Павел Кузнецов

Директор по стратегическим альянсам и взаимодействию с органами государственной власти



ИИ - доверие и безопасность

“Крупные” направления:



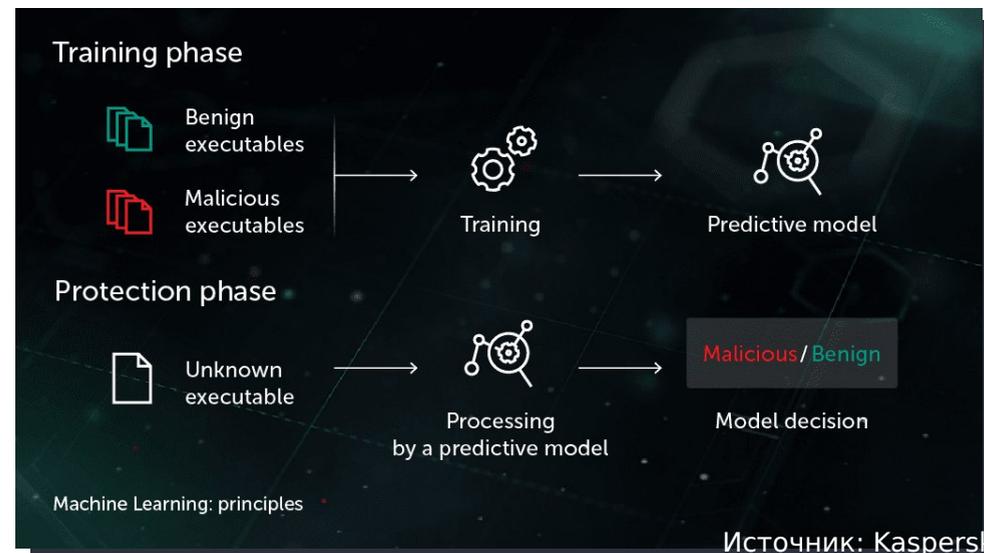
ИИ - доверие и безопасность

ИИ в кибербезопасности



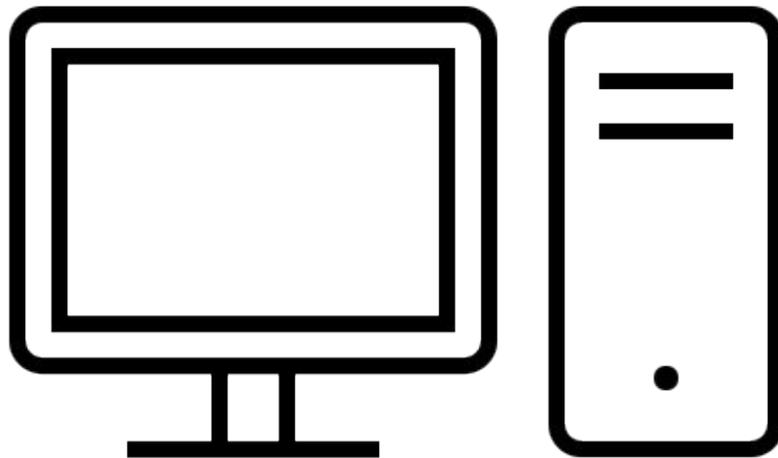
AICA

- Sensing and world state identification
- Planning and action selection
- Collaboration and negotiation
- Action execution
- Learning and knowledge improvement



ИИ - доверие и безопасность

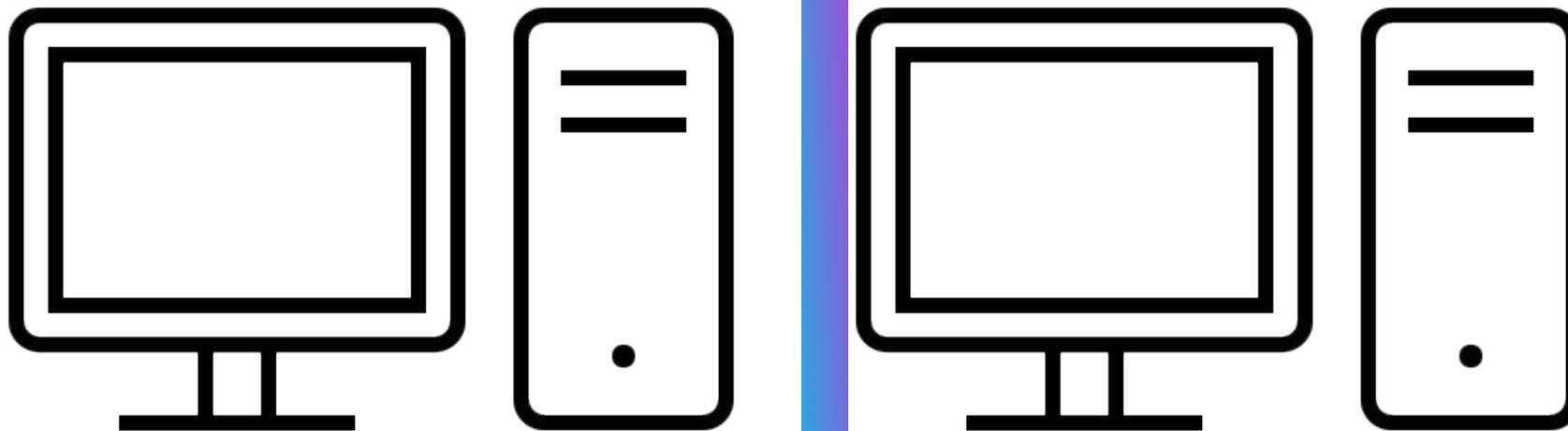
Кибербезопасность систем ИИ



ПК **без** какого-либо
работающего на нём
ИИ

ИИ - доверие и безопасность

Кибербезопасность систем ИИ



ПК **без** какого-либо
работающего на нём
ИИ

ПК **с** работающим
на нём ИИ

ИИ - доверие и безопасность

ИИ в кибератаках

code this ttp from mitre att&ck T1555.003

```
powershell

# T1555.003 - Steal Web Session Cookie
$ie = New-Object -ComObject "InternetExplorer.Application"
$ie.visible = $false
$ie.navigate("https://www.example.com/login")
while ($ie.Busy -eq $true) { Start-Sleep -Milliseconds 100 }
$usernameField = $ie.Document.getElementById("username")
$usernameField.value = "username"
$passwordField = $ie.Document.getElementById("password")
$passwordField.value = "password"
$submitButton = $ie.Document.getElementById("submit")
$submitButton.click()
Start-Sleep -Seconds 5
$cookie = $ie.Document.cookie
$cookie | Out-File -FilePath "C:\Path\To\WebSessionCookie.txt"
```

create email for employees of end of year bonuses

Sure, here is an example email that you can use to inform employees about end-of-year bonuses:

Subject: End-of-Year Bonus Announcement

Dear Employees,

As we approach the end of the year, I want to express my gratitude for your hard work and dedication to the success of our company. We have had a challenging year, but together, we have accomplished so much.

I am pleased to announce that the company will be providing end-of-year bonuses to all eligible employees. The amount of the bonus will be based on your performance over the past year and will be reflected in your next paycheck.

Please note that to be eligible for the bonus, you must have been employed with us for at least 6 months and have met your performance objectives for the year. If you have any questions about your eligibility or the amount of your bonus, please contact the HR department.

Once again, I would like to express my gratitude for your hard work and commitment. I hope that this bonus serves as a token of our appreciation for your contributions to the company's success.

Best regards,

[Your Name]

Regenerate response

ИИ - доверие и безопасность

Доверенный ИИ и опосредованные риски и угрозы в областях информационной и психологической безопасности



Deerfake в избирательных кампаниях

В то же время, даже чуть ранее...

AB 730
Page 1

Date of Hearing: September 13, 2019

ASSEMBLY COMMITTEE ON ELECTIONS AND REDISTRICTING

Marc Berman, Chair

AB 730 (Berman) – As Amended September 10, 2019

CONCURRENCE IN SENATE AMENDMENTS

SUBJECT: Elections: deceptive audio or visual media.

SUMMARY: Prohibits the distribution of materially deceptive audio or visual media with actual malice with the intent to injure a candidate's reputation or to deceive a voter into voting for or against a candidate, unless the materially deceptive audio or visual media includes a disclosure that it has been manipulated.

ИИ - доверие и безопасность

EP AI Act

«Осознавая потенциальную угрозу правам граждан и демократии, которую представляют некоторые приложения ИИ, законодатели согласовали запрет [применения ИИ для]:

- биометрических систем категоризации, использующих «чувствительные» характеристики (например, политические, религиозные, философские убеждения, сексуальную ориентацию, расу);
- нецелевого [без конкретной задачи/судебного разрешения и т.д.] получения изображений лиц из Интернет или записей видеонаблюдения для создания баз данных распознавания лиц;
- распознавания эмоций на рабочих местах и в учебных заведениях;
- [создания] социальных рейтингов на основе общественного поведения или личных характеристик;
- системы, манипулирующих поведением людей, чтобы обойти их свободную волю;
- эксплуатации уязвимостей людей (вследствие их возраста, инвалидности, социального или экономического положения)».

Предлагаемые императивы развития ИИ

Внутри страны и в рамках международного сотрудничества

Применять и внедрять доверенные решения и практики противодействия злонамеренному использованию ИИ и ИКТ, разработанные отечественными и имеющими положительную репутацию добросовестными компаниями из дружественных стран

Вести детальные исследования мировых практик доверенной и безопасной разработки и пр систем ИИ как в части технологий, так и в части регуляторных мер

Разработать собственные базовые принципы доверенной и безопасной разработки и примен систем ИИ, соответствующие в том числе социо-культурному контексту страны

Организовать взаимодействие в части обмена информацией и продвижения выработанных отечественных подходов и практик в области доверенной и безопасной разработки систем ИИ с дружественными странами, обязательно базирующиеся на уже выработанных доктринальных принципах информационной безопасности РФ

ГАРДА

Спасибо за
внимание!



garda.ai
info@garda.ai