

Субъектность ИИ с позиций КВАНТОВОЙ КОГНИТИВИСТИКИ

Суров Илья Алексеевич

Семинар «Этические проблемы искусственного интеллекта»
11 октября 2023 г., Москва, ИФ РАН

ИТМО

 **БЛАГОТВОРИТЕЛЬНЫЙ
ФОНД ВЛАДИМИРА
ПОТАНИНА**
СТИПЕНДИАЛЬНАЯ ПРОГРАММА

Университет ИТМО, г. Санкт-Петербург

Проблема: этически-правовой статус ИИ



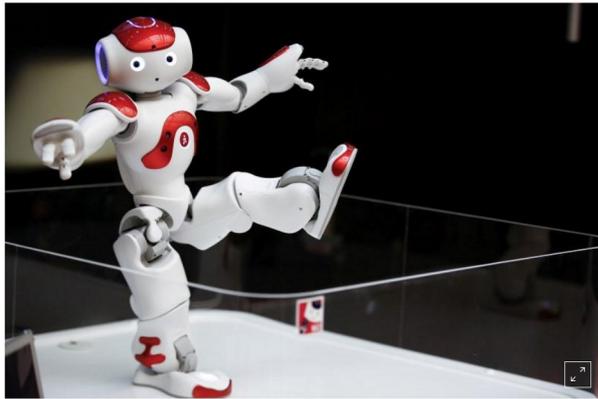
SCIENCE & SPACE JUNE 21, 2016 / 8:08 PM / UPDATED 7 YEARS AGO

Europe's robots to become 'electronic persons' under draft plan

By Georgina Prodhan

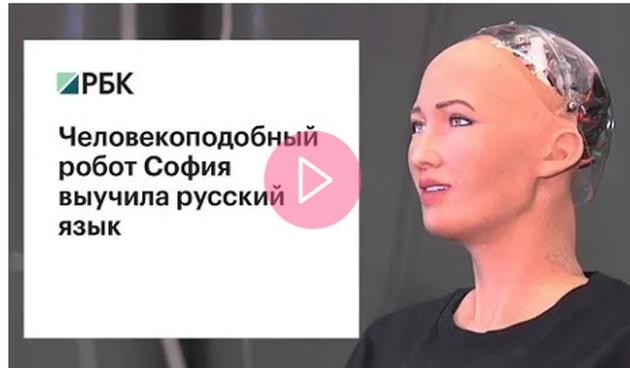
4 MIN READ

MUNICH, Germany (Reuters) - Europe's growing army of robot workers could be classed as "electronic persons" and their owners liable to paying social security for them if the European Union adopts a draft plan to address the realities of a new industrial revolution.



A "Nao" humanoid robot, by Aldebaran Robotics that offers basic service information, moves during a presentation at a branch of the Bank of Tokyo-Mitsubishi UFJ (MUFG) in Tokyo April 13, 2015. REUTERS/Thomas Peter/FILE Photo

AI 'boy' granted residency in central Tokyo



Robot granted Saudi citizenship has more rights than Saudi women



Кодекс этики в сфере ИИ
<https://ethics.a-ai.ru/>

https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html

UNESCO (2021).
The Ethics of Artificial Intelligence.

<https://www.reuters.com/article/us-europe-robotics-lawmaking-idUSKCN0Z72AY>

<https://phys.org/news/2017-11-ai-boy-granted-residency-central.html>

https://www.theregister.com/2017/10/27/robot_granted_saudi_citizenship/

Юридическая правосубъектность



Критерии:

- **Осведомлённость** (understanding):
Восприятие и запоминание фактической информации об инциденте
- **Понимание** (appreciation):
Понимание значения фактов и обстоятельств в приложении к проблеме, личностный смысл
- **Самостоятельность** (autonomy):
Отсутствие принуждения к действию, свобода принятия решений



По стандартным критериям правосубъектностью может быть наделён обычный робот-пылесос

Вопрос о природе ИИ?

Подобен человеку и
другим живым существам

Подобен другим машинам
и программам

Вопрос о природе человека!

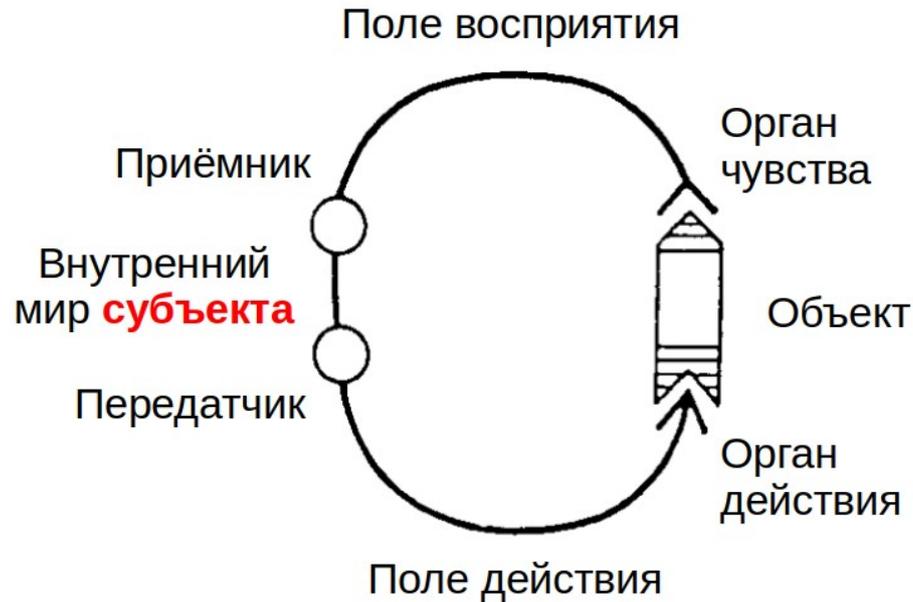
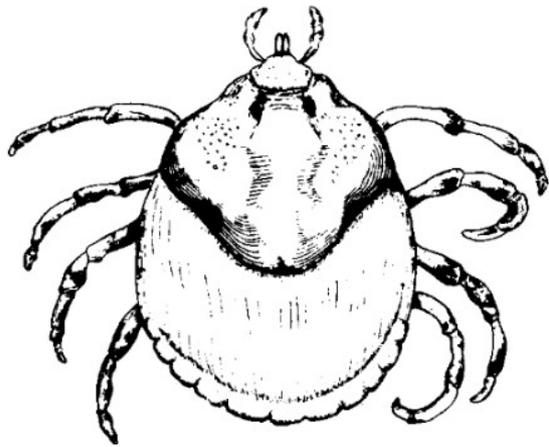
**Отличаемся ли мы от машин,
и если да, то чем?**

«Сверхвызов» современности.

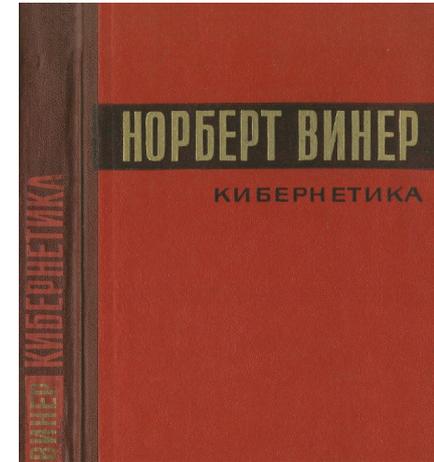
От игры в го до осознания себя: изменит ли искусственный разум мир людей

Переслегин С.Б. (2023) <https://fedpress.ru/article/3254098>

Что говорит наука?



Яков фон Иксюль
1864 - 1944



von Uexküll J. A stroll through the worlds of animals and men:
A picture book of invisible worlds // Semiotica. 1992. Vol. 89, № 4. P. 319–391.

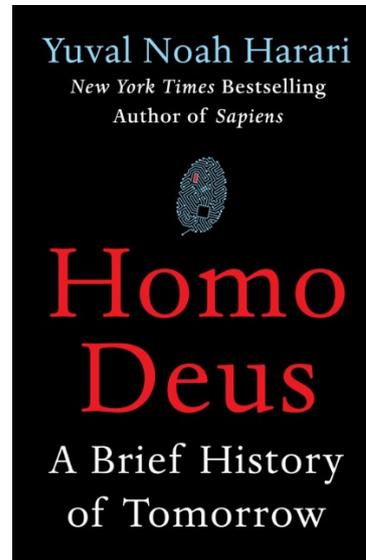
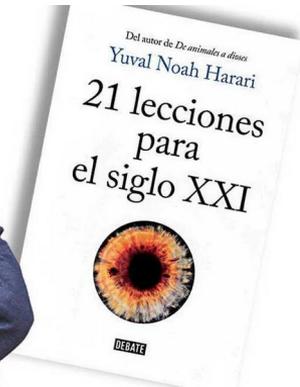
Брушлинский А.В. Психология субъекта / ред. Знаков В.В.
Санкт-Петербург: Алетейя, 2003. 272 р.

Винер Н. Кибернетика, или управление и связь в животном и машине.
// М.: Советское радио, 1968. 330 с.

Классический детерминизм

Абсолютная предопределённость будущего

Лаплас, Декарт, Эйнштейн



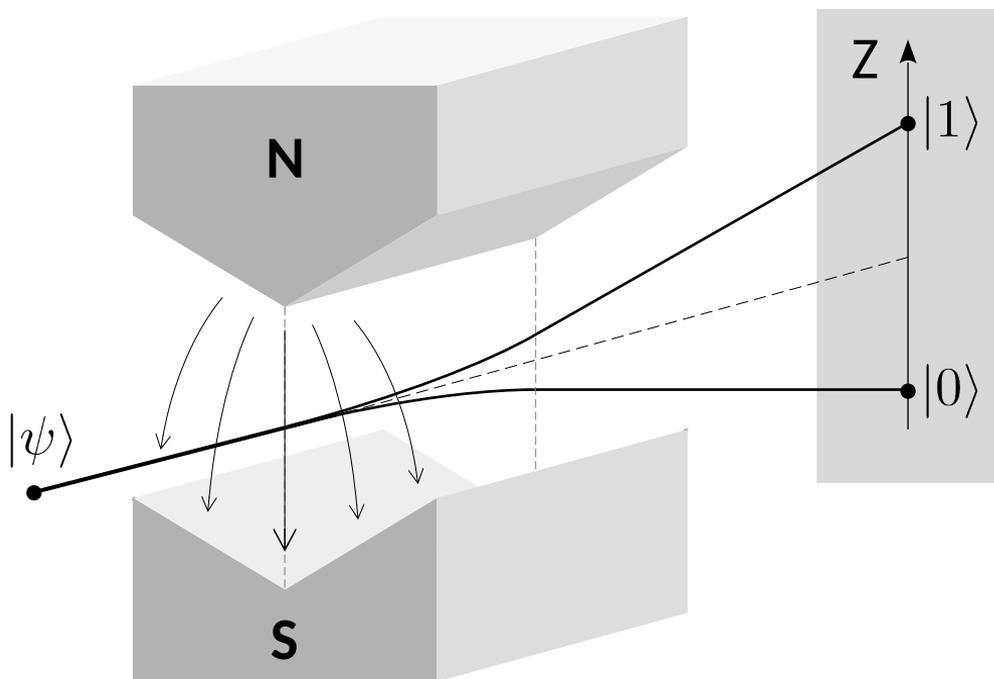
The liberal belief in individualism is founded on the three important assumptions that we discussed earlier in the book:

1. I am an in-dividual – i.e. I have a single essence which cannot be divided into any parts or subsystems. True, this inner core is wrapped in many outer layers. But if I make the effort to peel these external crusts, I will find deep within myself a clear and single inner voice, which is my authentic self.
2. My authentic self is completely free.
3. It follows from the first two assumptions that I can know things about myself nobody else can discover. For only I have access to my inner space of freedom, and only I can hear the whispers of my authentic self. This is why liberalism grants the individual so much authority. I cannot trust anyone else to make choices for me, because no one else can know who I really am, how I feel and what I want. This is why the voter knows best, why the customer is always right and why beauty is in the eye of the beholder.

However, the life sciences challenge all three assumptions. According to the life sciences:

1. Organisms are algorithms, and humans are not individuals – they are 'dividuals', i.e. humans are an assemblage of many different algorithms lacking a single inner voice or a single self.
2. The algorithms constituting a human are not free. They are shaped by genes and environmental pressures, and take decisions either deterministically or randomly – but not freely.
3. It follows that an external algorithm could theoretically know me much better than I can ever know myself. An algorithm that monitors each of the systems that comprise my body and my brain could know exactly who I am, how I feel and what I want. Once developed, such an algorithm could replace the voter, the customer and the beholder. Then the algorithm will know best, the algorithm will always be right, and beauty will be in the calculations of the algorithm.

Опыт Штерна — Герлаха



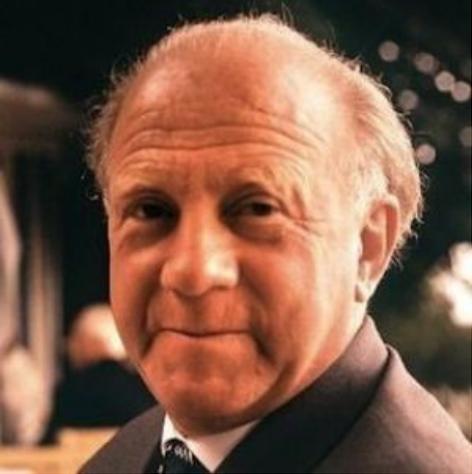
Квантовая случайность:

Исход не
предопределён никаким
законом природы.

Kofler J., Zeilinger A. Quantum information and randomness // Eur. Rev. 2010. Vol. 18, № 4. P. 469–480.

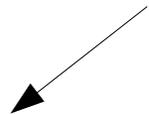
Jaeger G. Quantum randomness and unpredictability // Fortschritte der Phys. 2017. Vol. 65, № 6–8. P. 1600053.

Gerlach W., Stern O. Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld // Zeitschrift für Phys. 1922. Vol. 9, № 1. P. 349–352.



В. Гейзенберг

Неопределённость



Субъективная

Был ли дождь
вчера в Москве?

Объективная

Какая будет следующая
фраза в беседе?

Определение

Измерение

Решение

Результат

Предопределён

Непредопределён

Вероятность

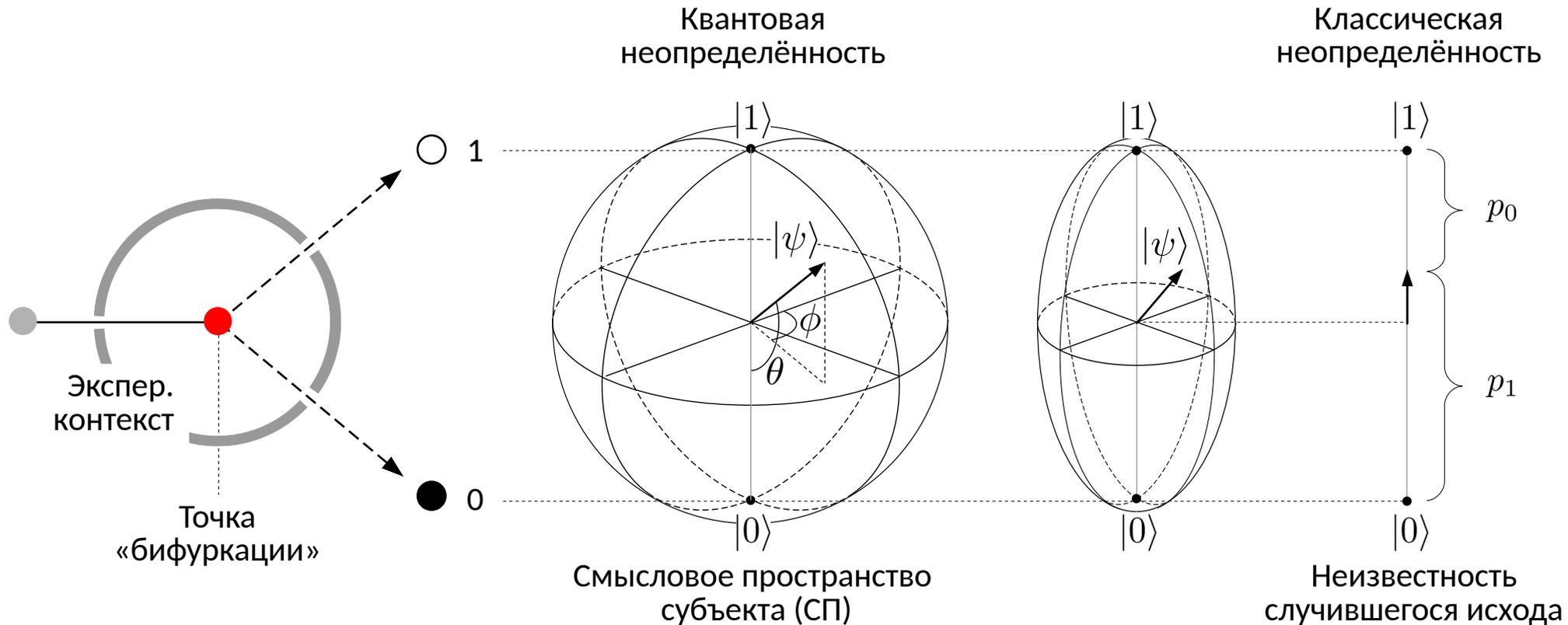
Классическая

Квантовая

Bell, J. S. (1990). **Against «measurement»**.
Physics World, 3, 32–41.

Гуревич И.И., Фейгенберг И.М. **Какие вероятности работают в психологии?** // Вероятностное прогнозирование деятельности человека. Москва: Наука, 1977. с. 9–21.

Математика



Двухвариантная
неопределённость

Кубит

Бит

Квантовая когнитивистика

Использование квантовой теории вероятности для моделирования мышления и поведения.

Гуревич И.И., Фейгенберг И.М. **Какие вероятности работают в психологии?** // Вероятностное прогнозирование деятельности человека. Москва: Наука, 1977. с. 9–21.

Орлов Ю.Ф. **A quantum model of doubt** // Ann. N. Y. Acad. Sci. 1981. Vol. 373. P. 84–92.

Aerts D. **Quantum structures: An attempt to explain the origin of their appearance in nature** // Int. J. Theor. Phys. 1995. Vol. 34, № 8. P. 1165–1186.

Хренников А.Ю. **Classical and quantum mechanics on information spaces with applications to cognitive, psychological, social, and anomalous phenomena** // Found. Phys. 1999. Vol. 29, № 7. P. 1065–1098.

Менский М.Б. **Концепция сознания в контексте квантовой механики** // Успехи физических наук. 2005. Т. 175, с. 413–435.



Хренников А.Ю.



Аертс Д.

Хренников А.Ю. **Quantum-like modeling of cognition** // Front. Phys. 2015. Vol. 3, № 77. P. 77.

Wendt A. **Quantum Mind and Social Science**. Cambridge University Press, 2015. 354 p.

Хренников А.Ю. **Ubiquitous Quantum Structure. From psychology to finance**. Heidelberg: Springer, 2010. 216 p.

Почему это работает?

Формальный подход:

Просто удобная математика.
«Shut up and calculate»

Суров И.А. **Какая разница? Прагматическая формализация смысла** // Искусственный интеллект и принятие решений. 2023. № 1.

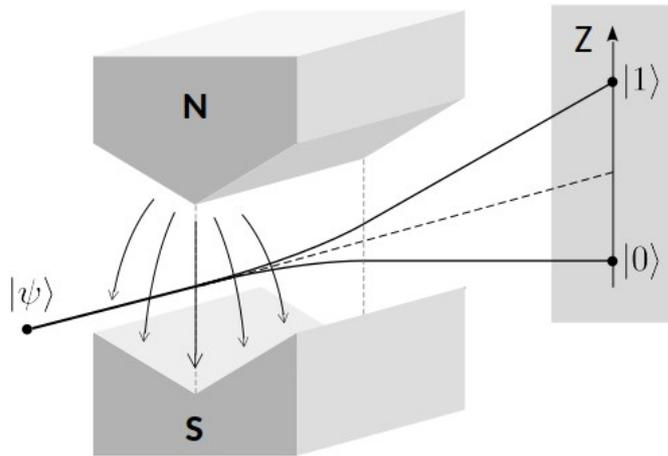
Surov I.A. **Natural Code of Subjective Experience** // Biosemiotics. 2022. Vol. 15, № 1.

Surov I.A. **Quantum core affect. Color-emotion structure of semantic atom** // Front. Psychol. 2022.

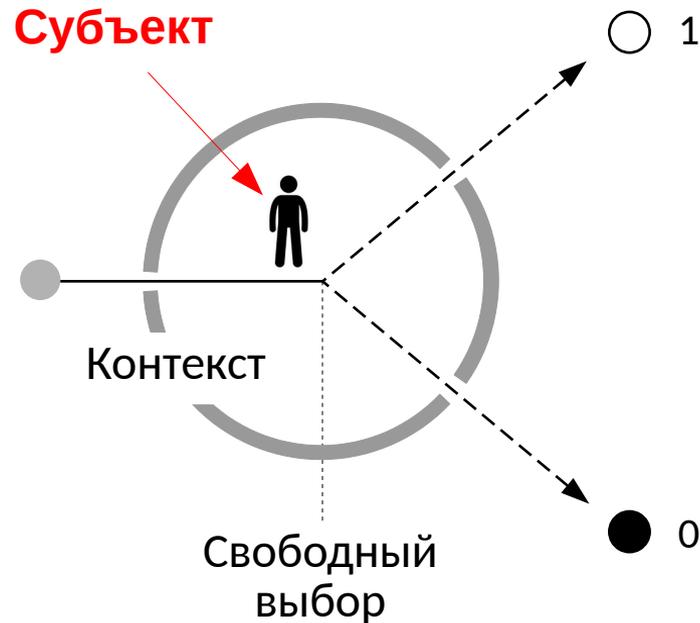
Сущностный подход

- Кубит кодирует «психологию», т.е. эмоционально-смысловое состояние субъекта.
- Живые организмы усиливают квантовую неопределённость с атомного до макроскопического масштаба.
- **Поведение человека не предопределено законами природы. Объективная квантовая неопределённость.**

Квантовая многовариантность — место для субъектности



Двухвариантное решение =
единица субъектного поведения



Stapp H.P.
**Quantum Theory
and Free Will.**
Springer, 2017. 142 p.

Wendt A.
**Quantum Mind and
Social Science.**
Cambridge University
Press, 2015. 354 p.

Объективная неопределённость

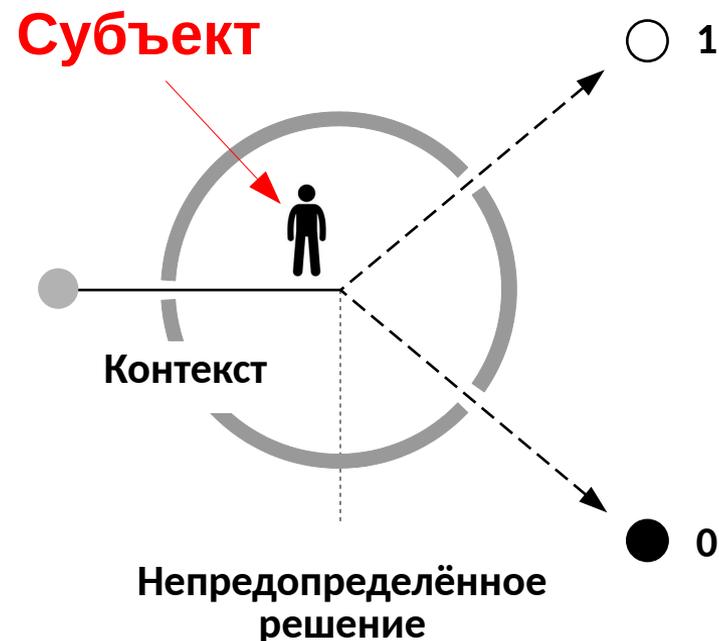


Васнецов В.М. 1882 г.

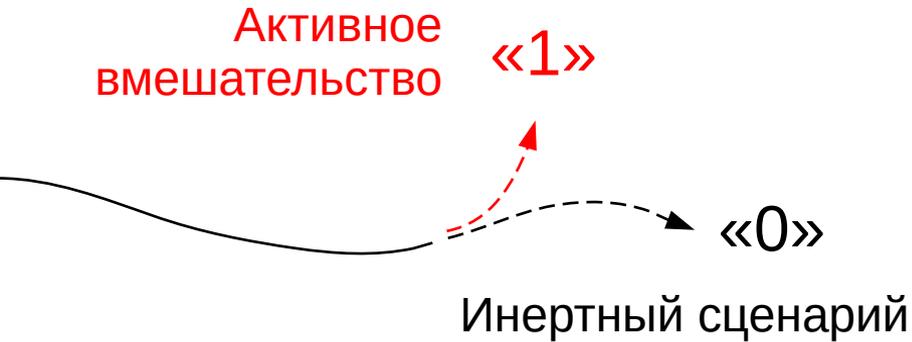
Определение субъекта:

Сущность,
различающая объективную
многовариантность будущего
и совершающая
соответствующий выбор.

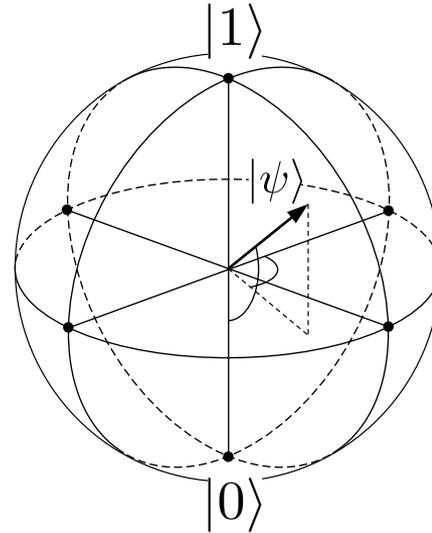
«ЛПР»



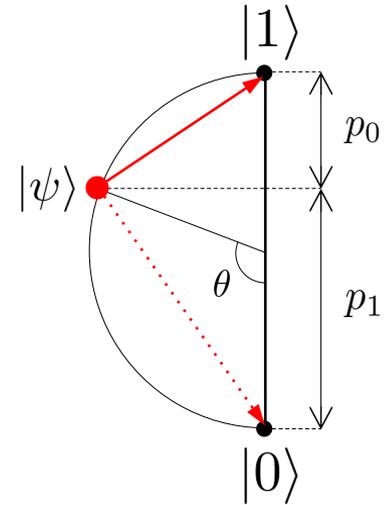
Компоненты субъектности



Сознание: Различение объективной возможности (квантовой неопределённости)



Мышление: Определение эмоционально-смыслового состояния $|\psi\rangle$



Воля: решение на основе $|\psi\rangle$

Классический ИИ

- | | | |
|--|--|--|
| 1) Различение объективных возможностей | нет, т.к. алгоритмически распознаются только заранее данные образы |  |
| 2) Определение квантового состояния | нет, т.к. пространство квантовых состояний не сформировано на шаге 1 | 
 |
| 3) Принятие решения | нет в силу отсутствия объективной возможности |  |

Юридическая правосубъектность

Те же самые критерии на основе квантовых понятий:

- **Осведомлённость** (understanding):
Восприятие и запоминание фактической информации об инциденте
- **Понимание** (appreciation):
Понимание значения фактов и обстоятельств в приложении к проблеме, **личностный смысл**
- **Самостоятельность** (autonomy):
Отсутствие принуждения к действию, **свобода принятия решений**



Выводы

В смысле (право) субъектности сколь угодно сложный классический ИИ неотличим от любого другого инструмента в силу отсутствия доступа к квантовой неопределённости;

Квантовый ИИ может обладать прото-субъектностью наравне с элементарными физическими системами;

Правовое регулирование и этические исследования ИИ упрощаются путём уточнения понятий

- принятия решений
- свободы выбора и воли
- субъективного смысла

на основе понятия о квантовой неопределённости.