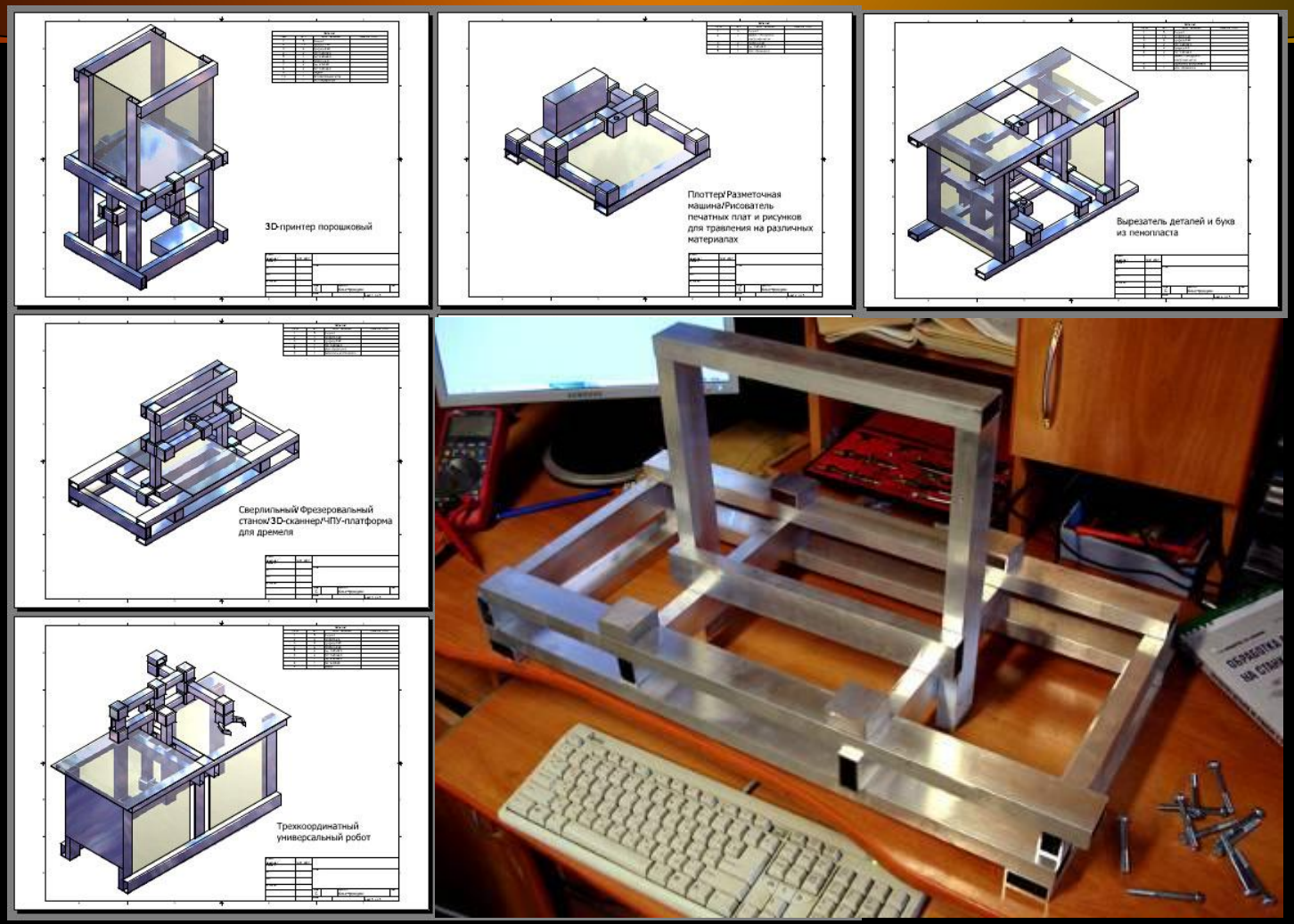
Семинар XX

ТРЕБУЮТСЯ ТГ АКТИВИСТЫ

- Фильм «Фонтан»
- **АКЦИЯ 1 МАРТА**

3D-принтер (Оликевич)



25 февраля, Моск. обл., д. Фенино, д. 108 (ехать по Рязанскому ш. до Бронниц или на электр. от м. Выхино до ст. Гжель, оттуда такси)

Ярославль (Батин)

Конгресс

РОССИЙСКОЕ
ФИЛОСОФСКОЕ ОБЩЕСТВО
119992, Москва, ул. Волхонка, 14
тел.: (495) 637-24-02
E-mail: [illegible]

RUSSIAN
PHILOSOPHICAL SOCIETY
14, Volhonka Str., Moscow, 119992
Russia. Tel.: (7-495) 637-24-02
Website: [illegible]

02.02.2007 № 77

Korean Organizing Committee.
Congress Secretariat WCP 2008 c/o ConvEx Korea
B-201 Champs-Elysees Center # 889-5, Daechi-dong, Gangnam-gu, Seoul 135-712, Korea
Tel: 822 - 557-77-05. Fax 822557-77-08, e-mail: wcp2008@convex.co.kr

PROPOSAL FOR A SYMPOSIUM

Russian Philosophical Society ask you to consider our proposal for holding a symposium "Immortalism and transhumanism: philosophical study of new technological determinant of societal development" during the XXII World Congress on Philosophy.

This initiative has been endorsed by a number of international organisations. Letters expressing their support are enclosed together with this proposal.

February 2, 2007

President of the Russian Philosophical Society,
Director of science, Institute of Philosophy of the Russian Academy of Sciences,
Head of sector of social sciences, Russian Academy of Sciences

Vyacheslav Stepin

ПРЕЗИДИУМ

Технологические аспекты СИНГУЛЯРНОСТИ

Медведев Данила

Вглядываясь в Сингулярность

Винж, 1982	**Винж, 1993**	**Курцвейл, 200**
умнее человека	закон Мура	связь с ускорение
темпы не важны	2030	ускорение прогресс

- Определения *«нельзя предсказать»*
  «прогресс вертикальный»
  «создание сверхразума»

- Математика
  ускорение
  → графики
  → «предсказание»

- Причинный анализ
  - причины непредсказуемости
  - характер изменений

Непредсказуемость

Неопределённость
(самомодификация)

Взрывной рост
(исчерпание трендов за конечный срок)

Ускоряющийся рост

Возникновение сверхинтеллекта
(непредсказуемость для человека)

Причины непредсказуемости

непредсказуемость сегодня

- умнее человека
- мы не можем предсказать, что будет после С
- проблема с моделями
- мы не можем предсказать поведение более умной сущности, чем мы
- иначе мы могли бы вести себя так же

Винж, 1982

неопределённость

- самомодификация
- правила не постоянны
- в ИИ изначально есть доступ к коду
- ИИ и нано
- два направления, которые самоусиливаются
- возможно, математическая непредсказуемость
- бифуркации

предсказуемость!

- если известна цель
- Юдковски, 2006
- но действия всё равно не известны
- антропоморфизм - это не решение
- т. к. ИИ очень разные
- математическое решение
- Рефлексивная теория принятия решений (понимание работы сознания, переписывающего свой код)

Проблемы с математикой

- математика предсказывает
- причинный анализ объясняет
- качественный анализ описывает следствия

- математика не важна для кач. модели
  - порог: переход через чел. интеллект
  - темпы роста могут расти, могут снижаться
- вопрос: верно ли мат. модели описывают динамику?
  - на разных этапах
  - нет гарантии, что тенденции сохранятся

Сценарии

- Четыре пути
  - ИИ
    - создание ИИ
    - поумнение сетей
  - УИ
    - интерфейсы мозг-компьютер
    - биологическое усиление

В. Винж, 1993

Когнитивная наука

- Радикальное усиление интеллекта

- ПО для организации знания, управления ими, генерации новых знаний
- Оптимизация биологического мозга
  - ноотропики
  - прямая стимуляция (TMS)
  - доращивание (стволовые клетки)
- интеграция
  - интерфейсы мозг-компьютер (I/O)
  - полная интеграция
  - загрузка

Нанотехнологии

- Нанотех – это инструмент для контроля над физическим миром
  - а не полумеры вроде роботов

- Нанокомпьютеры могут быть нужны для сверхмощного ИИ
- Сложность нанотехнологий может требовать ИИ

- Достаточно ли оптимизации кода для взрывного роста ИИ?

ДА

- создаётся ИИ
- Сингулярность
- риск ИИ
- создаются нанотехнологии
- риск нанотеха

НЕТ

- создаётся ИИ
- разрабатываются нанотехнологии
- риск нанотеха
- усиливается ИИ
- Сингулярность
- риск ИИ

Когда ждать Сингулярности?

- ИИ-группы
  - Singularity Institute
  - AGIRI (Novamente)
  - a2i2 (Adaptive A.I.)
  - Ai Research
- Загрузка
  - Brains in Silicon (Stanford)
  - CCortex (Artificial Development)
  - Blue Brain (IBM)

- Сроки реализации проектов?
- Ожидаемые пути развития событий?