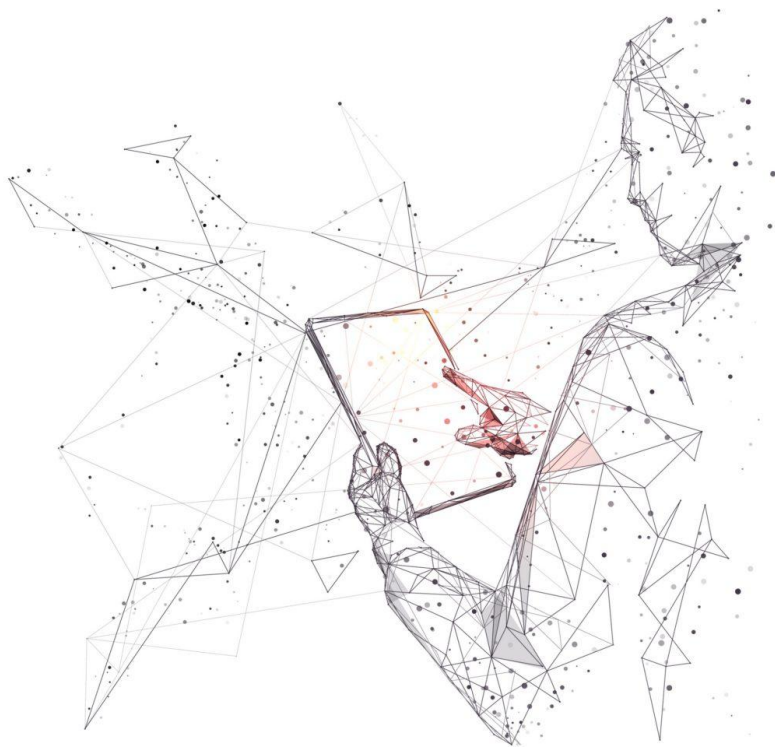


Международная научная конференция XVI Академические чтения, посвященные 100-летию академика РААСН Бондаренко В.М. и 96-летию академика РААСН Осипова Г.Л.
«АКТУАЛЬНЫЕ ВОПРОСЫ СТРОИТЕЛЬСТВА. НАДЕЖНОСТЬ СТРОИТЕЛЬНЫХ КОНСТРУКЦИЙ. ЭНЕРГОСБЕРЕЖЕНИЕ. ЭКОЛОГИЧЕСКАЯ БЕЗОПАСНОСТЬ. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»



КАРТЫ ЗНАНИЙ:

**на пути к доверенной
информационной среде
коллективного
человеко-машинного разума**

*Воронцов Константин
Вячеславович*

д.ф.-м.н., профессор РАН
(МГУ, МФТИ, ФИЦ ИУ РАН)

Большие языковые модели (Large Language Models, LLM)



уже встраиваются
в бизнес-процессы через

поисковые и рекомендательные системы

системы документооборота

системы видеоконференцсвязи (RConf.ru)

корпоративные базы знаний



будут всё шире
встраиваться в процессы



оказания
помощи

юридической

медицинской

психологической



образования, воспитания



исследований и научного поиска

в роли интерфейса между
человеком и знаниями



как обеспечить доверенность в коммуникации с машиной на естественном языке?

Аспекты доверенности больших языковых моделей (LLM)



технические,
**относительно
системы**

качество, эффективность, надёжность

робастность, устойчивость к изменениям условий среды

безопасность, устойчивость к атакам любого вида



неприемлемость
для человека

галлюцинаций

токсичности

обмана, манипуляций



психологические,
**относительно
человека**



предотвращение
у человека

когнитивных искажений

неоправданных ожиданий

мизинтерпретаций



эргономичность,
комфортность

способа коммуникации

адаптации к пользователю

Почему люди доверяют друг другу? Применимо ли это к LLM?



Проблемы на пути к доверенным LLM



проблема 1:
несовершенство (пока)
самых языковых моделей

обучающих данных

нейросетевых архитектур

методик оценивания



проблема 2:
неэффективность текста

как средства передачи знаний

в условиях переизбытка информации



проблема 3:
дефицит консенсуса
относительно

целей, задач, критериев, приоритетов

способов представления знаний

способов взаимодействия людей с LLM



проблема 4:
недооценённость визуального
аналитического мышления

как способа представления знаний

как инструмента выработки консенсуса

Проблема №1: несовершенство LLM



несовершенство (пока) обучающих данных

терабайтов текста, накопленных человечеством

избыточность

противоречивость

неструктурированность



несовершенство (пока) архитектуры моделей

обучаемых предсказывать слова по контексту

не обладающих человеческими

структурами мышления

критериями важности

картинами мира

проявляющееся в ошибках, «галлюцинациях»

несовершенство (пока) методик оценивания, недостаточность тестов, бенчмарков



решается путём

отбора, фильтрации обучающих данных

расширения многокритериальных бенчмарков

обучения по хорошо структурированным текстам

как именно структурированным? откуда их взять? в больших объёмах!

Проблема №2: неэффективность текста

текст из драйвера
развития цивилизации
становится его тормозом

как средства передачи знаний от головы к голове

из-за затрат на кодирование знаний в текст и декодирование их из текста обратно



при том, что в головах
знания структурированы

иерархически, по важности

образно, нечётко, неточно

частично, у всех по-разному



избыточность текста
может скрывать

нелогичность, неточность, неполноту

манипуляции, заблуждения, обман

флуд, демагогию, графоманию



решается
путём

постепенного отказа от линейных текстов в пользу

нелинейного чтения (non-linear reading)

«сначала главное, детализация по требованию»

технологий текстографического представления знаний

Проблема №3: дефицит консенсуса

относительно системы ценностей, целей, задач, критериев, приоритетов



относительно способа
представления знаний,
компромиссного для

индивидуального восприятия человеком

коллективной интеллектуальной деятельности

машинной обработки

взаимодействия человека и машины (LLM)



относительно *доверенного*
источника знаний с единой
структурой и критериями

Интернет? — много мусора

Википедия? — нет полноты и единства

наука? — дефицит структурированности



создания глобальной доверенной **Системы Знаний**,
на основе *радиантной текстографической структуры*



решается
путём

- сначала вручную, затем полуавтоматически с помощью LLM
- сначала в проектах, затем в отраслях, в странах, в мире

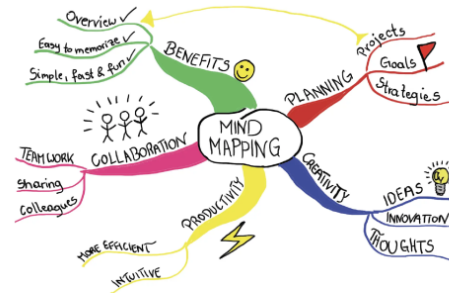


- дополнения **Системы Знаний** механизмами
- коммуникации, выработки консенсуса
 - интеграции с бизнес-процессами

Интеллект-карты (mind maps) Тони Бьюзена



способ визуализации
того, как темы (мысли, идеи)
разбиваются на подтемы
иерархически



благодаря активизации обоих полушарий мозга и учёта особенностей восприятия, мышления, памяти



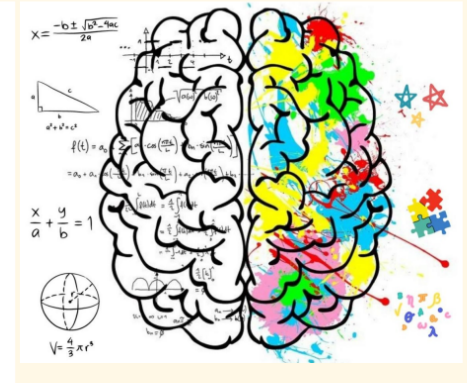
нацелены на
повышение
эффективности

понимания

запоминания

систематизации

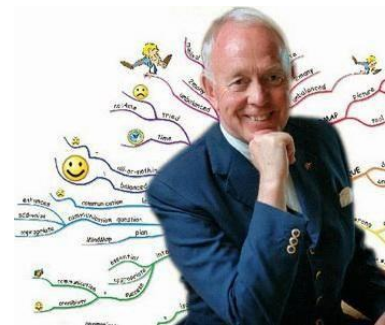
поиска консенсуса



техника
запоминания

посмотреть, понять, обсудить, принять

самостоятельно воспроизвести через
10 минут → сутки → неделю → месяц



Интеллект-карты (mind-maps)



16 принципов построения интеллект-карт



графическое
оформление

активация
зрительной памяти

радиантность: линии расходятся из центра

размер шрифта отражает важность тем и подтем

цвет выделяет поддеревья

картинки усиливают образность

дополнение связями, выносками, ссылками



ветвления

однородность:

подтемы образуют нарратив, сюжет

либо отвечают на общий вопрос

полнота: подтемы охватывают все аспекты темы

точность: среди подтем невозможно выделить лишнюю

компактность: у темы 5 ± 3 подтем (число Ингве-Миллера 7 ± 2)

значимость: (под)темы отбираются и ранжируются по важности

16 принципов построения интеллект-карт



эргономичности

наглядность: слова подкрепляются изображениями

лаконичность: темы формулируются максимально кратко

обозримость: карту понимают и запоминают целиком



эстетичности

красота, живость: эмоциональные карты лучше запоминаются

гармоничность: впечатление целостности, сложности карты

сбалансированность: ветви примерно равны и равноценны

КАРТЫ ЗНАНИЙ усиливают mind-map (6) новыми принципами



(1)

отторгаемость

комментарии автора не обязательны для понимания карты

карта способна «жить своей жизнью»

компромисс с лаконичностью



(2)

читабельность

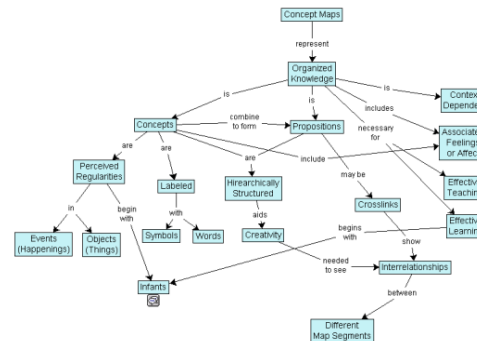
любой фрагмент карты
читается как связный текст

легко и однозначно

даже автоматически

в отличие от
других техник
представления
знаний
— онтологий
— фреймов и др.

компромисс с
лаконичностью
и обозримостью



КАРТЫ ЗНАНИЙ усиливают mind-map (6) новыми принципами



(3)
сворачиваемость

компромисс
с читабельностью

любой темы без утраты

читабельности

сбалансированности

под любой формат: книги, постера, экрана, презентации

чтобы любую детализацию «отложить на потом»

способствует

выделению главного в каждой теме

пониманию и взаимопониманию



(4)
коллективность,
на всех этапах
жизненного цикла

компромиссы между авторами

создание

рецензирование, согласование

развитие

уточнение, реструктуризация

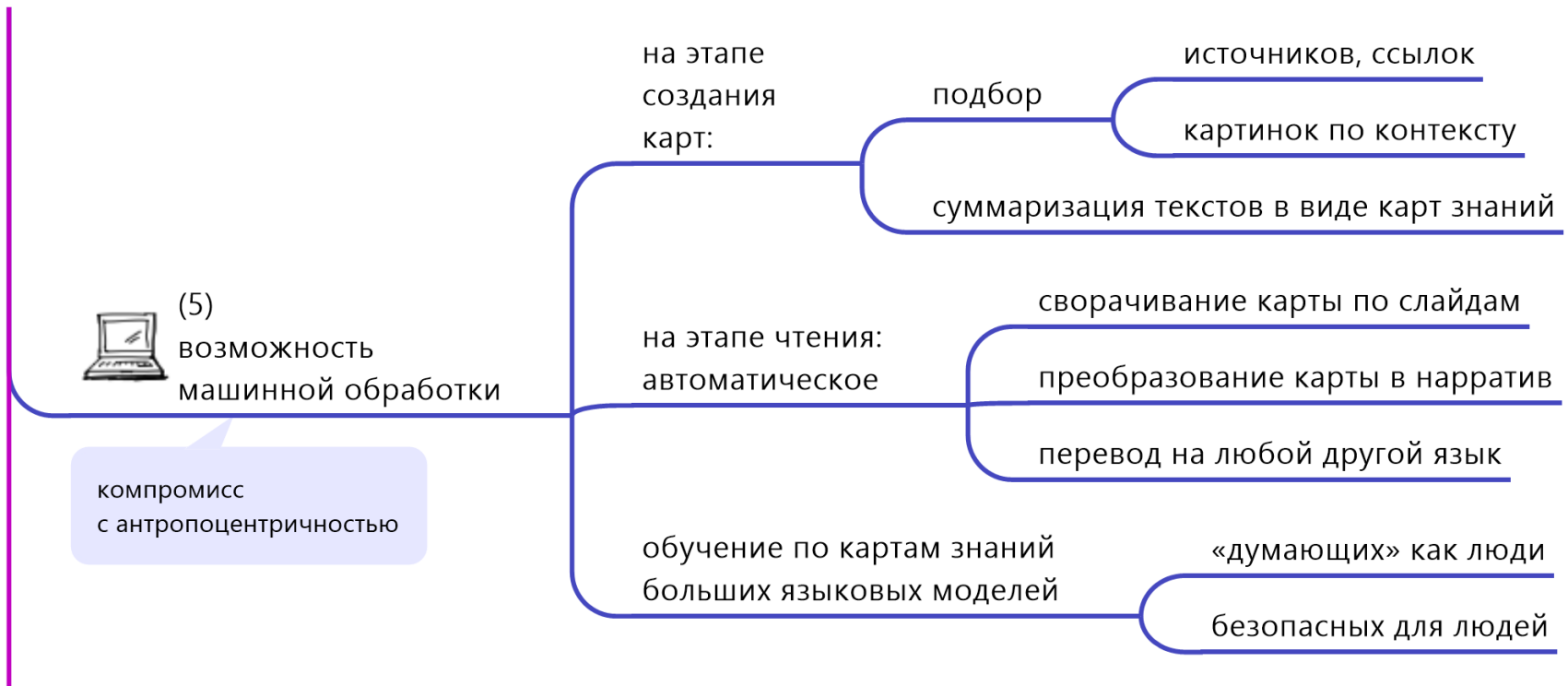
детализация, разрастание

применение

в практической деятельности

с разграничением прав доступа

КАРТЫ ЗНАНИЙ усиливают mind-map (6) новыми принципами



КАРТЫ ЗНАНИЙ усиливают mind-map (6) новыми принципами



Смысловое ядро глобальной Системы Знаний



Как включить визуальное аналитическое мышление



решается
путём

активации и практики, **в четыре этапа:**

1 порядка сотни карт: просмотреть, обсудить, поспорить, принять

2 десятки карт: построить самому, следуя 11+6 принципам

3 испытать инсайты,
«моменты ясности»,
когда карта



индивидуальная практика и опыт

«красиво сложилась»

привела к согласию

легко и ярко запомнилась,

легла в основу деятельности

4 сделать построение карт регулярной
профессиональной практикой



индивидуальной

коллективной

ВЫВОДЫ. Карты знаний:

основаны на интеллект-картах (mind-map) Тони Бьюзена

отличаются более строгими принципами построения (11+6)

активируют **визуальное аналитическое мышление**

ведут к взаимопониманию, практическому **коллективному разуму**

удобно (оптимально?) представляют знания для человека и машины



способствуют взаимопониманию и доверию между людьми, значит, способны также обеспечить доверие людей к LLM, если LLM будет общаться на языке **радиантных структур**

строиться эффективнее при автоматизации средствами AI / LLM

обеспечить создание единой Системы Знаний как среды для практической деятельности коллективного разума

обеспечить реструктуризацию и верификацию знаний для обучения доверенных LLM следующего поколения

внедрить в AI человеческую цивилизационную систему ценностей

обеспечить доверенность в **человеко-машинной цивилизации**

могут

в порядке
усиления
гипотез



КАРТЫ ЗНАНИЙ:

на пути к доверенной информационной среде коллективного
человеко-машинного разума



Воронцов Константин Вячеславович

k.vorontsov@iai.msu.ru

д.ф.-м.н., профессор РАН
рук. лаб. машинного обучения и семантического анализа ИИИ МГУ,
зав. каф. математических методов прогнозирования ВМК МГУ,
зав. каф. машинного обучения и цифровой гуманитаристики МФТИ