
Человек и искусственный интеллект: ПОДХОДЫ К ВЗАИМОДЕЙСТВИЮ

Д.В. Ушаков

Взаимодействие человека и ИИ

- Цифровой ангел.
- Франкенштейн.



Индивидуальный цифровой ангел

- Помощь в организации себя.
 - Помощь во взаимодействии с людьми.
 - Профессиональная поддержка.
 - Защита от обмана.
-

Психологические состояния

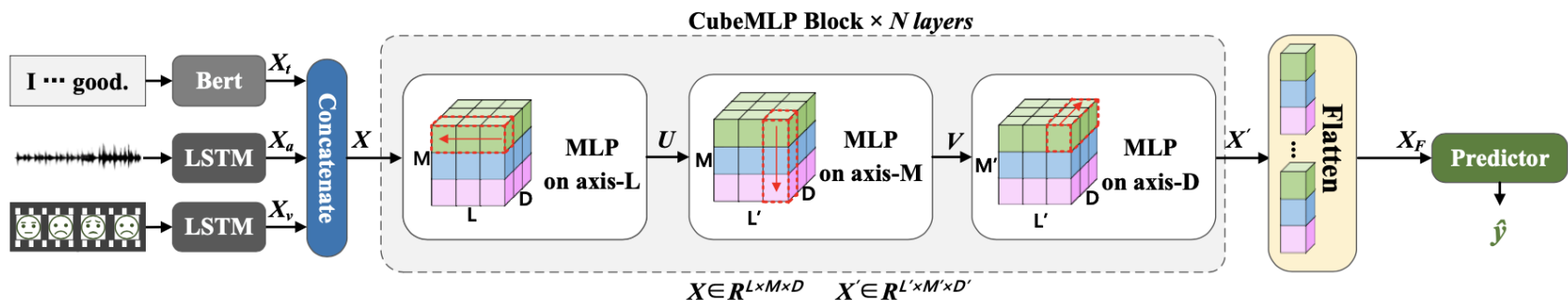
- Труд в сложных условиях: транспорт, авиация, АЭС, работа вахтовым методом, космонавтика.
 - Образование.
 - Персонал компаний.
 - Психологическое состояние населения.
 - Переговоры, продажи.
 - Эмоциональный ИИ.
-

Существующие подходы

- **Источники данных**
 - Видеозаписи.
 - Психофизиологические показатели (ЭКГ, ЭЭГ, МРТ).
 - **Разметка данных**
 - Экспертная оценка.
 - Самооценка.
 - Оценочные процедуры (тестирование).
-

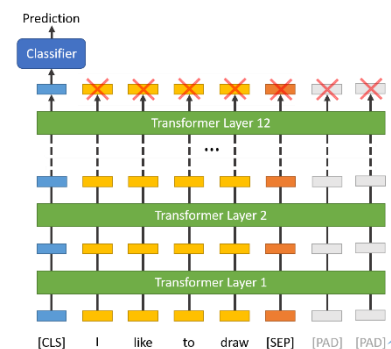
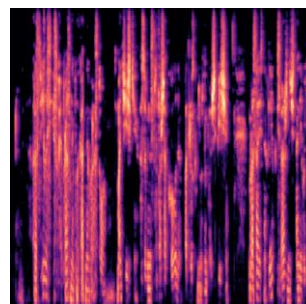
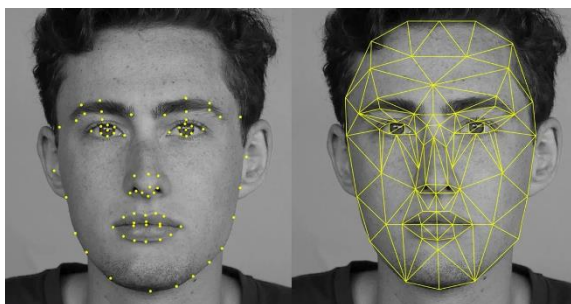
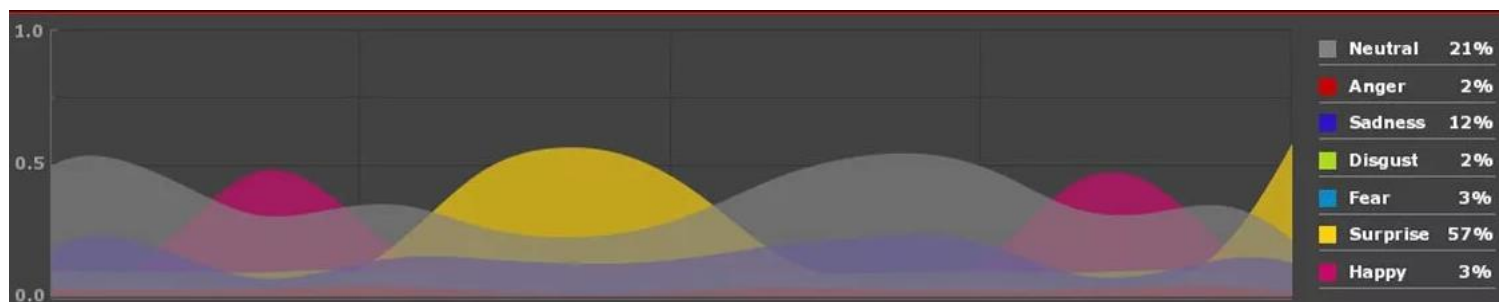
Объединение данных

- Объединение данных разных модальностей.
- Обучение с учителем на базе различных архитектур нейросетей.



Существующий подход

- Определение психологических свойств личности по видео-интервью.

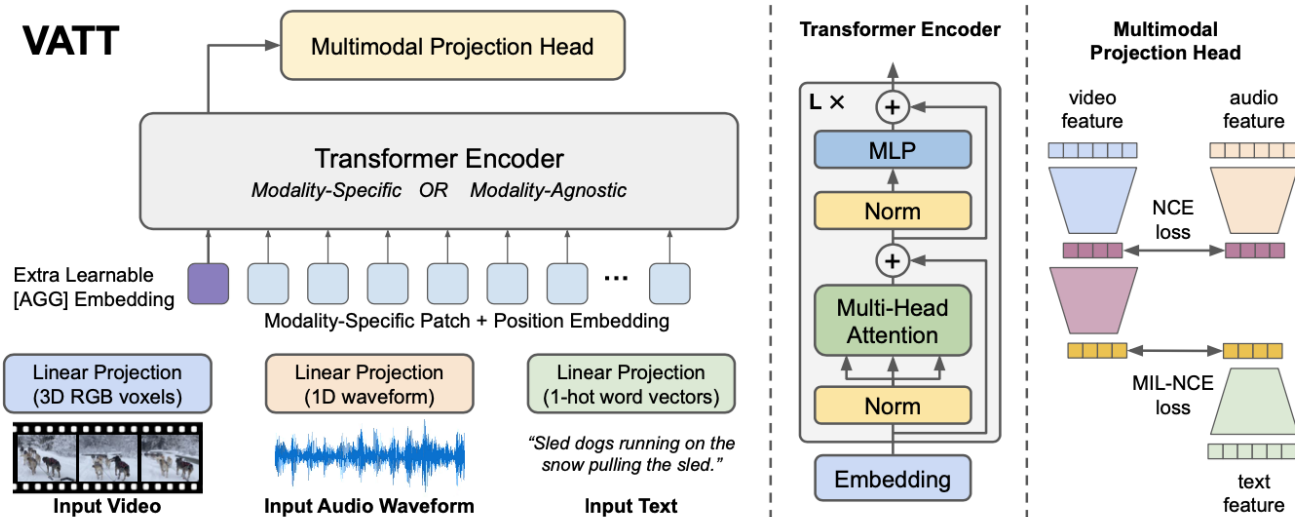


Недостатки существующих ПОДХОДОВ

- Ограничение пространства состояний.
 - Оторванность оценки состояния от порождающей его ситуации.
 - «Узкая» специализация созданных моделей при высокой трудоемкости процедуры обучения и сбора данных.
-

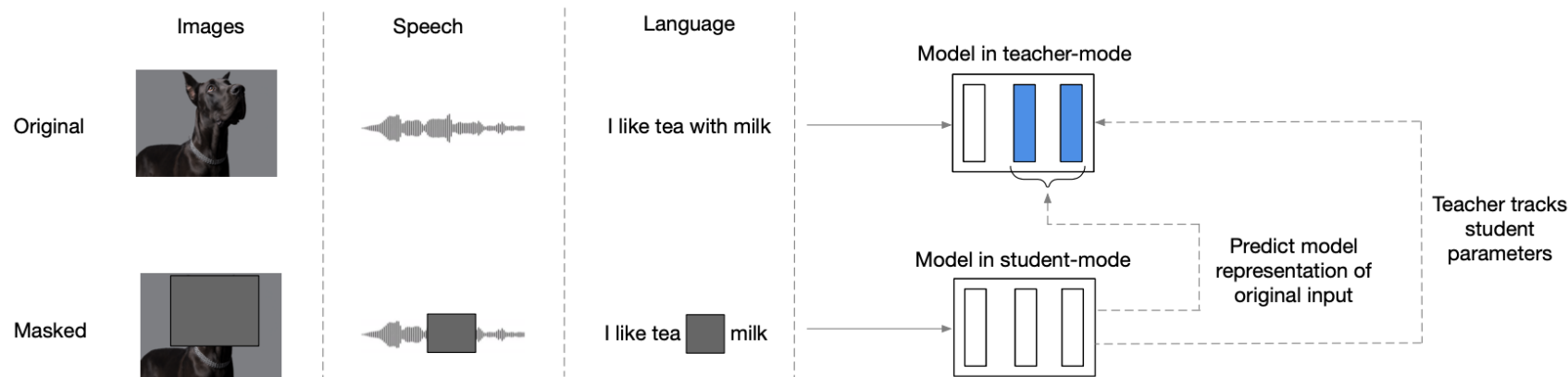
Ближайшая перспектива

- Обучение без учителя (Self-Supervised Learning) на базе архитектуры трансформера.
- Модель создает репрезентацию объектов в новом пространстве – «эмбединг».



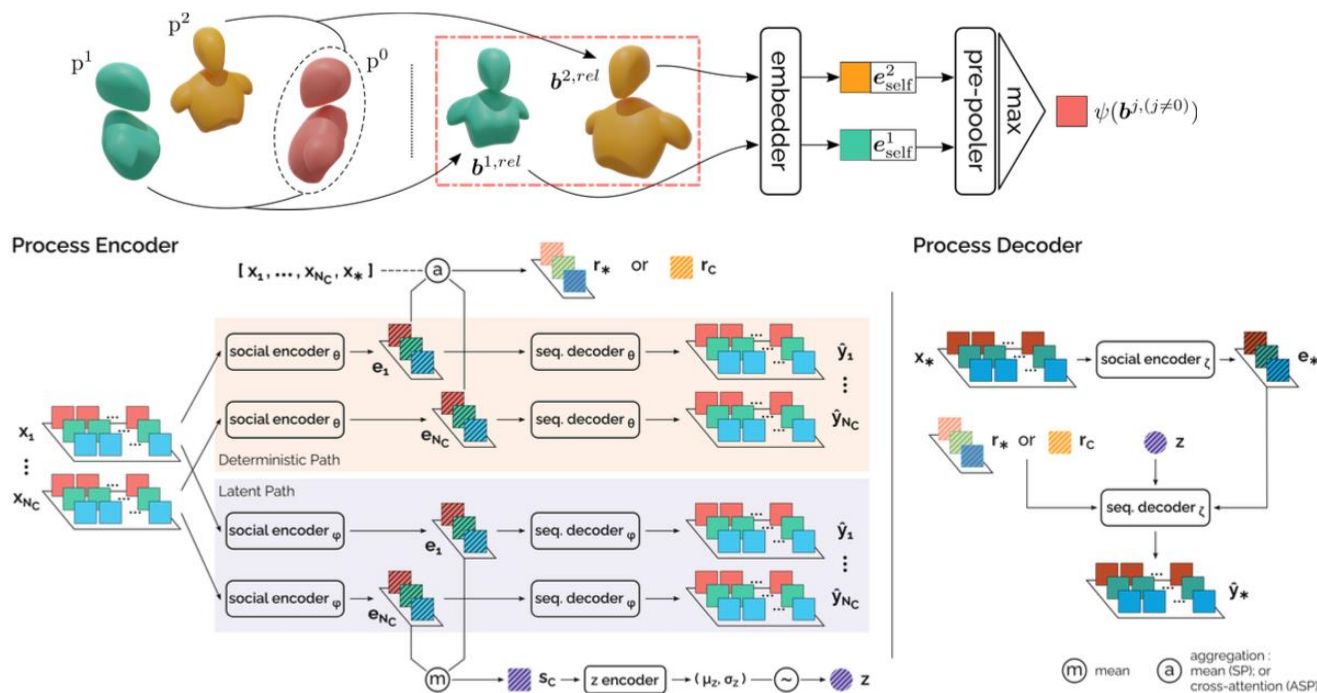
Ближайшая перспектива

- Обучение на базе трансформеров не требует размеченной выборки для всего объема данных.
- Осуществляется через маскирование данных.



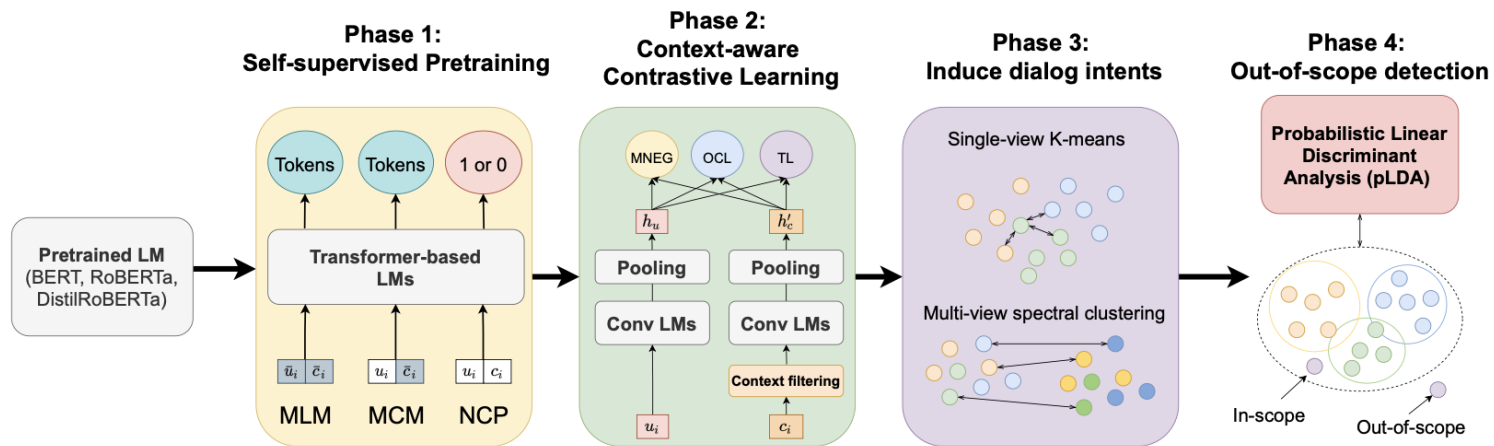
Ближайшая перспектива

- Обучение на записях взаимодействия людей с целью прогноза следующего состояния.



Ближайшая перспектива

- Определение интенций по тексту в процессе диалогов с учетом контекста.



Преимущества

- Сведение данных различных модальностей в единое пространство эмбедингов.
 - Построение апостериорного пространства психологических состояний человека.
 - Создание моделей определения и предсказания состояний людей в процессе взаимодействия.
 - Отсутствие необходимости разметки всей обучающей выборки при возможности использования всего многообразия доступных данных.
-

Применение

- Образование.
- Напряженная деятельность.
- Состояние коллектива или людей в местах скоплений.
- Эмоциональный искусственный интеллект.
- Оценка состояния участников совещания в реальном времени.
- Продажи.
- Call-центры.

Индивидуальный цифровой ангел

- Помощь в организации себя.
 - Помощь во взаимодействии с людьми.
 - Профессиональная поддержка.
 - Защита от обмана.
-

Помощь в организации себя

- Регулировка состояния, борьба со стрессом, депрессией и т.д.
 - Создание индивидуальных образовательных траекторий.
 - Консультация в выборе профессии и др. с учетом индивидуальных особенностей.
 - Организация режима дня.
-

Помощь во взаимодействии с ЛЮДЬМИ

- Отношения с близкими людьми, в семье.
 - Отношения с детьми.
 - Отношения в профессиональной сфере.
 - Взаимодействие между «цифровыми ангелами».
-

Профессиональная поддержка

- Проведение переговоров.
- Продажи.



Защита от обмана

- Оценка добросовестности собеседника.
 - Оценка информации, поступающей из СМИ.
-

Франкенштейн

- Образ муравейника.
 - Интенциональность искусственного интеллекта.
-

Интенциональность ИИ

- Возможность управления физической энергией.
 - Наличие независимых мотивации и целеполагания.
-

Независимость мотивации ИИ

- От человечества в целом.
- Зависимость от групп людей.



ИИ и группы людей

- Модель ответственности государств была отработана в ситуации ядерного противостояния.
 - В ситуации искусственного интеллекта ответственность смещается на уровень корпораций и групп людей, способных разрабатывать и воплощать технологии искусственного интеллекта.
 - Техно-гуманитарный баланс.
-

Независимость от человечества

- Техническая ошибка.
 - Необходимость или целесообразность.
 - Не надо ли допустить искусственный интеллект к определению того, что мы хотим?
-

Чего мы хотим?

- Инстинкты у животных – результат естественного отбора на выживание.
- Ген и мем.
- Ценностные системы людей подвержены быстрому изменению.
- Индивидуальное равновесие ценностей.
- Социальное равновесие ценностей.
- Конкуренция групп с разными ценностями и изменение условий жизни.

Наша мотивационная система

- До сих пор позволяет нам развиваться и не допустила катастрофы.
 - Несовершенна, нет доказательств ее безопасности, постоянно ведет к конфликтам и жертвам.
 - Вмешательство искусственного интеллекта: повышение благосостояния, совершенствование решений.
 - Человек – момент жизни разума.
-

Благодарю за внимание!
