

# СИНЕРГЕТИЧЕСКОЕ ОБЪЯСНЕНИЕ ПОЛЕЗНОСТИ ЦЕЛЕНАПРАВЛЕННОГО ПОВЕДЕНИЯ

Смолин В.С.

Институт прикладной математики им. М.В. Келдыша РАН, РФ,  
125047, Москва, Миусская пл., д.4, +7 499 978-13-14, smolin@keldysh.ru

Рассматриваются проблемы создания общего искусственного интеллекта (AGI) на основе нейросетевых архитектур в сложных средах. Успех современных нейросетей (включая большие языковые модели, LLM) связан не только с ростом вычислительной мощности, но и со способностью соответствовать закону необходимого разнообразия [1]. Но для достижения AGI недостаточно простого масштабирования параметров.

Важно научить AGI декомпозиции сложной среды на простые, статистически изучаемые компоненты. Человек легко производит такое разбиение, что отражено, например, в структуре языка, которую используют LLM. Однако в нейросетевых теориях эта проблема не рассматривается как центральная [2, 3]. AGI должен уметь осуществлять декомпозицию и строить иерархии простых моделей.

Вторая проблема — комбинаторный взрыв вариантов действий. Даже обладая моделями компонентов среды, перебор всех вариантов взаимодействий невозможен. В новых комбинациях могут эмерджентно возникать непредсказуемые свойства. Решение — в локальности взаимодействий и моделировании развития только реально возникших и ожидаемых ситуаций при достижении поставленных целей.

Выделим два режима использования знаний: 1) Интуитивный: быстрое выполнение действий на основе хорошо изученных процессов без моделирования последствий; 2) Мышление: моделирование вариантов действий и их последствий на основе знаний о компонентах для новых, нестандартных ситуаций.

Критически важным является контроль переключения между этими режимами на основе оценки потенциальных рисков. Этот механизм управления распределением вычислительных ресурсов между быстрым «интуитивным» реагированием и ресурсоёмким «мышлением» можно назвать «нейросетевым сознанием».

Путь к AGI лежит через создание архитектур, автономно осуществляющих декомпозиции, использующих иерархические модели для формирования целей, и управляющих своими ресурсами под контролем нейросетевого сознания.

## Литература.

1. У. Р. Эшби. Введение в кибернетику. — М.: Иностранная литература, 1959.
2. Чуприкова Н. И. Психика и психические процессы (Система понятий общей психологии). — М. : Языки славянской культуры : Знак, 2015. — 608 с.
3. R. L. Kuhn. A landscape of consciousness: Toward a taxonomy of explanations and implications. *Progress in Biophysics and Molecular Biology*, Volume 190, August 2024, Elsevier, Pages 28-169. <https://doi.org/10.1016/j.pbiomolbio.2023.12.003>
4. Малинецкий Г.Г., Смолин В.С. Биологическое сознание – интуиция и мышление без использования логики // Искусственный интеллект. Теория и практика, 2023, №12(2), с. 2-14.